(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/46426 A2

(51) International Patent Classification[7]: C12N 15/52, 9/00, A01K 67/027, C07K 16/40, C12Q 1/68, G01N 33/573, 33/577, 33/53, A61K 38/43, 39/395

(21) International Application Number: PCT/US01/47429

(22) International Filing Date: 4 December 2001 (04.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/254,308	8 December 2000 (08.12.2000)	US
60/256,189	15 December 2000 (15.12.2000)	US
60/257,713	21 December 2000 (21.12.2000)	US
60/262,706	19 January 2001 (19.01.2001)	US
60/266,020	2 February 2001 (02.02.2001)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SANJANWALA, Madhu, M.** [US/US]; 210 Sylvia Court, Los Altos, CA 94024 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **ARVIZU, Chandra** [US/US]; 490 Sherwood Way, #1, Menlo Park, CA 94025 (US). **RING, Huijun, Z.** [US/US]; 625 Orange Avenue, Los Altos, CA 94022 (US). **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **DING, Li** [CN/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **WARREN, Bridget, A.** [US/US]; 10130 Parwood Drive #2, Cupertino, CA 95014 (US). **YANG, Junming** [CN/US]; 7125 Bank Lane, San Jose, CA 95129 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street, #205, San Leandro, CA 94577 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/46426 A2

(54) Title: DRUG METABOLIZING ENZYMES

(57) Abstract: The invention provides human drug metabolizing enzymes (DME) and polynucleotides which identify and encode DME. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of DME.

DRUG METABOLIZING ENZYMES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of drug metabolizing enzymes and to the use of these sequences in the diagnosis, treatment, and prevention of autoimmune/inflammatory, cell proliferative, neurological, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of drug metabolizing enzymes.

BACKGROUND OF THE INVENTION

The metabolism of a drug and its movement through the body (pharmacokinetics) are important in determining its effects, toxicity, and interactions with other drugs. The three processes governing pharmacokinetics are the absorption of the drug, distribution to various tissues, and elimination of drug metabolites. These processes are intimately coupled to drug metabolism, since a variety of metabolic modifications alter most of the physicochemical and pharmacological properties of drugs, including solubility, binding to receptors, and excretion rates. The metabolic pathways which modify drugs also accept a variety of naturally occurring substrates such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins. The enzymes in these pathways are therefore important sites of biochemical and pharmacological interaction between natural compounds, drugs, carcinogens, mutagens, and xenobiotics.

It has long been appreciated that inherited differences in drug metabolism lead to drastically different levels of drug efficacy and toxicity among individuals. For drugs with narrow therapeutic indices, or drugs which require bioactivation (such as codeine), these polymorphisms can be critical. Moreover, promising new drugs are frequently eliminated in clinical trials based on toxicities which may only affect a segment of the patient group. Advances in pharmacogenomics research, of which drug metabolizing enzymes constitute an important part, are promising to expand the tools and information that can be brought to bear on questions of drug efficacy and toxicity (See Evans, W.E. and R.V. Relling (1999) Science 286:487-491).

Drug metabolic reactions are categorized as Phase I, which functionalize the drug molecule and prepare it for further metabolism, and Phase II, which are conjugative. In general, Phase I reaction products are partially or fully inactive, and Phase II reaction products are the chief excreted species. However, Phase I reaction products are sometimes more active than the original administered drugs; this metabolic activation principle is exploited by pro-drugs (e.g. L-dopa). Additionally, some nontoxic compounds (e.g. aflatoxin, benzo[*a*]pyrene) are metabolized to toxic

intermediates through these pathways. Phase I reactions are usually rate-limiting in drug metabolism. Prior exposure to the compound, or other compounds, can induce the expression of Phase I enzymes however, and thereby increase substrate flux through the metabolic pathways. (See Klaassen, C.D. et al. (1996) Casarett and Doull's Toxicology: The Basic Science of Poisons, McGraw-Hill, New York, NY, pp. 113-186; Katzung, B.G. (1995) Basic and Clinical Pharmacology, Appleton and Lange, Norwalk, CT, pp. 48-59; Gibson, G.G. and P. Skett (1994) Introduction to Drug Metabolism, Blackie Academic and Professional, London.)

Drug metabolizing enzymes (DMEs) have broad substrate specificities. This can be contrasted to the immune system, where a large and diverse population of antibodies are highly specific for their antigens. The ability of DMEs to metabolize a wide variety of molecules creates the potential for drug interactions at the level of metabolism. For example, the induction of a DME by one compound may affect the metabolism of another compound by the enzyme.

DMEs have been classified according to the type of reaction they catalyze and the cofactors involved. The major classes of Phase I enzymes include, but are not limited to, cytochrome P450 and flavin-containing monooxygenase. Other enzyme classes involved in Phase I-type catalytic cycles and reactions include, but are not limited to, NADPH cytochrome P450 reductase (CPR), the microsomal cytochrome b5/NADH cytochrome b5 reductase system, the ferredoxin/ferredoxin reductase redox pair, aldo/keto reductases, and alcohol dehydrogenases. The major classes of Phase II enzymes include, but are not limited to, UDP glucuronyltransferase, sulfotransferase, glutathione S-transferase, N-acyltransferase, and N-acetyl transferase.

Cytochrome P450 and P450 catalytic cycle-associated enzymes

Members of the cytochrome P450 superfamily of enzymes catalyze the oxidative metabolism of a variety of substrates, including natural compounds such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins, as well as drugs, carcinogens, mutagens, and xenobiotics. Cytochromes P450, also known as P450 heme-thiolate proteins, usually act as terminal oxidases in multi-component electron transfer chains, called P450-containing monooxygenase systems. Specific reactions catalyzed include hydroxylation, epoxidation, N-oxidation, sulfooxidation, N-, S-, and O-dealkylations, desulfation, deamination, and reduction of azo, nitro, and N-oxide groups. These reactions are involved in steroidogenesis of glucocorticoids, cortisols, estrogens, and androgens in animals; insecticide resistance in insects; herbicide resistance and flower coloring in plants; and environmental bioremediation by microorganisms. Cytochrome P450 actions on drugs, carcinogens, mutagens, and xenobiotics can result in detoxification or in conversion of the substance to a more toxic product. Cytochromes P450 are abundant in the liver, but also occur in other tissues; the enzymes are located in microsomes. (See ExPASY ENZYME EC 1.14.14.1; Prosite PDOC00081 Cytochrome P450 cysteine heme-iron ligand signature; PRINTS EP450I E-Class P450 Group I

signature; Graham-Lorence, S. and Peterson, J.A. (1996) FASEB J. 10:206-214.)

Four hundred cytochromes P450 have been identified in diverse organisms including bacteria, fungi, plants, and animals (Graham-Lorence, supra). The B-class is found in prokaryotes and fungi, while the E-class is found in bacteria, plants, insects, vertebrates, and mammals. Five subclasses or groups are found within the larger family of E-class cytochromes P450 (PRINTS EP450I E-Class P450 Group I signature).

All cytochromes P450 use a heme cofactor and share structural attributes. Most cytochromes P450 are 400 to 530 amino acids in length. The secondary structure of the enzyme is about 70% alpha-helical and about 22% beta-sheet. The region around the heme-binding site in the C-terminal part of the protein is conserved among cytochromes P450. A ten amino acid signature sequence in this heme-iron ligand region has been identified which includes a conserved cysteine involved in binding the heme iron in the fifth coordination site. In eukaryotic cytochromes P450, a membrane-spanning region is usually found in the first 15-20 amino acids of the protein, generally consisting of approximately 15 hydrophobic residues followed by a positively charged residue. (See Prosite PDOC00081, supra; Graham-Lorence, supra.)

Cytochrome P450 enzymes are involved in cell proliferation and development. The enzymes have roles in chemical mutagenesis and carcinogenesis by metabolizing chemicals to reactive intermediates that form adducts with DNA (Nebert, D.W. and Gonzalez, F.J. (1987) Ann. Rev. Biochem. 56:945-993). These adducts can cause nucleotide changes and DNA rearrangements that lead to oncogenesis. Cytochrome P450 expression in liver and other tissues is induced by xenobiotics such as polycyclic aromatic hydrocarbons, peroxisomal proliferators, phenobarbital, and the glucocorticoid dexamethasone (Dogra, S.C. et al. (1998) Clin. Exp. Pharmacol. Physiol. 25:1-9). A cytochrome P450 protein may participate in eye development as mutations in the P450 gene CYP1B1 cause primary congenital glaucoma (Online Mendelian Inheritance in Man (OMIM) *601771 Cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1; CYP1B1).

Cytochromes P450 are associated with inflammation and infection. Hepatic cytochrome P450 activities are profoundly affected by various infections and inflammatory stimuli, some of which are suppressed and some induced (Morgan, E.T. (1997) Drug Metab. Rev. 29:1129-1188). Effects observed in vivo can be mimicked by proinflammatory cytokines and interferons. Autoantibodies to two cytochrome P450 proteins were found in patients with autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), a polyglandular autoimmune syndrome (OMIM *240300 Autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy).

Mutations in cytochromes P450 have been linked to metabolic disorders, including congenital adrenal hyperplasia, the most common adrenal disorder of infancy and childhood; pseudovitamin D-deficiency rickets; cerebrotendinous xanthomatosis, a lipid storage disease characterized by

progressive neurologic dysfunction, premature atherosclerosis, and cataracts; and an inherited resistance to the anticoagulant drugs coumarin and warfarin (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc. New York, NY, pp. 1968-1970; Takeyama, K. et al. (1997) Science 277:1827-1830; Kitanaka, S. et al. (1998) N. Engl. J. Med. 338:653-661; OMIM *213700 Cerebrotendinous xanthomatosis; and OMIM #122700 Coumarin resistance). Extremely high levels of expression of the cytochrome P450 protein aromatase were found in a fibrolamellar hepatocellular carcinoma from a boy with severe gynecomastia (feminization) (Agarwal, V.R. (1998) J. Clin. Endocrinol. Metab. 83:1797-1800).

The cytochrome P450 catalytic cycle is completed through reduction of cytochrome P450 by NADPH cytochrome P450 reductase (CPR). Another microsomal electron transport system consisting of cytochrome b5 and NADPH cytochrome b5 reductase has been widely viewed as a minor contributor of electrons to the cytochrome P450 catalytic cycle. However, a recent report by Lamb, D.C. et al. (1999; FEBS Lett. 462:283-288) identifies a Candida albicans cytochrome P450 (CYP51) which can be efficiently reduced and supported by the microsomal cytochrome b5/NADPH cytochrome b5 reductase system. Therefore, there are likely many cytochromes P450 which are supported by this alternative electron donor system.

Cytochrome b5 reductase is also responsible for the reduction of oxidized hemoglobin (methemoglobin, or ferrihemoglobin, which is unable to carry oxygen) to the active hemoglobin (ferrohemoglobin) in red blood cells. Methemoglobinemia results when there is a high level of oxidant drugs or an abnormal hemoglobin (hemoglobin M) which is not efficiently reduced. Methemoglobinemia can also result from a hereditary deficiency in red cell cytochrome b5 reductase (Reviewed in Mansour, A. and Lurie, A.A. (1993) Am. J. Hematol. 42:7-12).

Members of the cytochrome P450 family are also closely associated with vitamin D synthesis and catabolism. Vitamin D exists as two biologically equivalent prohormones, ergocalciferol (vitamin D_2), produced in plant tissues, and cholecalciferol (vitamin D_3), produced in animal tissues. The latter form, cholecalciferol, is formed upon the exposure of 7-dehydrocholesterol to near ultraviolet light (i.e., 290-310 nm), normally resulting from even minimal periods of skin exposure to sunlight (reviewed in Miller, W.L. and Portale, A.A. (2000) Trends Endocrinol. Metab. 11:315-319).

Both prohormone forms are further metabolized in the liver to 25-hydroxyvitamin D (25(OH)D) by the enzyme 25-hydroxylase. 25(OH)D is the most abundant precursor form of vitamin D which must be further metabolized in the kidney to the active form, 1α,25-dihydroxyvitamin D (1α,25$(OH)_2$D), by the enzyme 25-hydroxyvitamin D 1α-hydroxylase (1α-hydroxylase). Regulation of 1α,25$(OH)_2$D production is primarily at this final step in the synthetic pathway. The activity of 1α-hydroxylase depends upon several physiological factors including the circulating level of the enzyme product (1α,25$(OH)_2$D) and the levels of parathyroid hormone (PTH), calcitonin, insulin,

calcium, phosphorus, growth hormone, and prolactin. Furthermore, extrarenal 1α-hydroxylase activity has been reported, suggesting that tissue-specific, local regulation of 1α,25(OH)2D production may also be biologically important. The catalysis of $1\alpha,25(OH)_2D$ to 24,25-dihydroxyvitamin D ($24,25(OH)_2D$), involving the enzyme 25-hydroxyvitamin D 24-hydroxylase (24-hydroxylase), also occurs in the kidney. 24-hydroxylase can also use 25(OH)D as a substrate (Shinki, T. et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94:12920-12925; Miller, W.L. and Portale, A.A. supra; and references within).

Vitamin D 25-hydroxylase, 1α-hydroxylase, and 24-hydroxylase are all NADPH-dependent, type I (mitochondrial) cytochrome P450 enzymes that show a high degree of homology with other members of the family. Vitamin D 25-hydroxylase also shows a broad substrate specificity and may also perform 26-hydroxylation of bile acid intermediates and 25, 26, and 27-hydroxylation of cholesterol (Dilworth, F.J. et al. (1995) J. Biol. Chem. 270:16766-16774; Miller, W.L. and Portale, A.A. supra; and references within).

The active form of vitamin D ($1\alpha,25(OH)_2D$) is involved in calcium and phosphate homeostasis and promotes the differentiation of myeloid and skin cells. Vitamin D deficiency resulting from deficiencies in the enzymes involved in vitamin D metabolism (e.g., 1α-hydroxylase) causes hypocalcemia, hypophosphatemia, and vitamin D-dependent (sensitive) rickets, a disease characterized by loss of bone density and distinctive clinical features, including bandy or bow leggedness accompanied by a waddling gait. Deficiencies in vitamin D 25-hydroxylase cause cerebrotendinous xanthomatosis, a lipid-storage disease characterized by the deposition of cholesterol and cholestanol in the Achilles' tendons, brain, lungs, and many other tissues. The disease presents with progressive neurologic dysfunction, including postpubescent cerebellar ataxia, atherosclerosis, and cataracts. Vitamin D 25-hydroxylase deficiency does not result in rickets, suggesting the existence of alternative pathways for the synthesis of 25(OH)D (Griffin, J.E. and Zerwekh, J.E. (1983) J. Clin. Invest. 72:1190-1199; Gamblin, G.T. et al. (1985) J. Clin. Invest. 75:954-960; and Miller, W.L. and Portale, A.A. supra).

Ferredoxin and ferredoxin reductase are electron transport accessory proteins which support at least one human cytochrome P450 species, cytochrome P450c27 encoded by the CYP27 gene (Dilworth, F.J. et al. (1996) Biochem. J. 320:267-71). A Streptomyces griseus cytochrome P450, CYP104D1, was heterologously expressed in E. coli and found to be reduced by the endogenous ferredoxin and ferredoxin reductase enzymes (Taylor, M. et al. (1999) Biochem. Biophys. Res. Commun. 263:838-42), suggesting that many cytochrome P450 species may be supported by the ferredoxin/ferredoxin reductase pair. Ferredoxin reductase has also been found in a model drug metabolism system to reduce actinomycin D, an antitumor antibiotic, to a reactive free radical species (Flitter, W.D. and Mason, R.P. (1988) Arch. Biochem. Biophys. 267:632-639).

Flavin-containing monooxygenase (FMO)

Flavin-containing monooxygenases oxidize the nucleophilic nitrogen, sulfur, and phosphorus heteroatom of an exceptional range of substrates. Like cytochromes P450, FMOs are microsomal and use NADPH and O_2; there is also a great deal of substrate overlap with cytochromes P450. The tissue distribution of FMOs includes liver, kidney, and lung.

There are five different known isoforms of FMO in mammals (FMO1, FMO2, FMO3, FMO4, and FMO5), which are expressed in a tissue-specific manner. The isoforms differ in their substrate specificities and other properties such as inhibition by various compounds and stereospecificity of reaction. FMOs have a 13 amino acid signature sequence, the components of which span the N-terminal two-thirds of the sequences and include the FAD binding region and the FATGY motif which has been found in many N-hydroxylating enzymes (Stehr, M. et al. (1998) Trends Biochem. Sci. 23:56-57; PRINTS FMOXYGENASE Flavin-containing monooxygenase signature).

Specific reactions include oxidation of nucleophilic tertiary amines to N-oxides, secondary amines to hydroxylamines and nitrones, primary amines to hydroxylamines and oximes, and sulfur-containing compounds and phosphines to S- and P-oxides. Hydrazines, iodides, selenides, and boron-containing compounds are also substrates. Although FMOs appear similar to cytochromes P450 in their chemistry, they can generally be distinguished from cytochromes P450 in vitro based on, for example, the higher heat lability of FMOs and the nonionic detergent sensitivity of cytochromes P450; however, use of these properties in identification is complicated by further variation among FMO isoforms with respect to thermal stability and detergent sensitivity.

FMOs play important roles in the metabolism of several drugs and xenobiotics. FMO (FMO3 in liver) is predominantly responsible for metabolizing (S)-nicotine to (S)-nicotine N-1'-oxide, which is excreted in urine. FMO is also involved in S-oxygenation of cimetidine, an H_2-antagonist widely used for the treatment of gastric ulcers. Liver-expressed forms of FMO are not under the same regulatory control as cytochrome P450. In rats, for example, phenobarbital treatment leads to the induction of cytochrome P450, but the repression of FMO1.

Endogenous substrates of FMO include cysteamine, which is oxidized to the disulfide, cystamine, and trimethylamine (TMA), which is metabolized to trimethylamine N-oxide. TMA smells like rotting fish, and mutations in the FMO3 isoform lead to large amounts of the malodorous free amine being excreted in sweat, urine, and breath. These symptoms have led to the designation fish-odor syndrome (OMIM 602079 Trimethylaminuria).

Lysyl oxidase

Lysyl oxidase (lysine 6-oxidase, LO) is a copper-dependent amine oxidase involved in the formation of connective tissue matrices by crosslinking collagen and elastin. LO is secreted as an N-glycosylated precursor protein of approximately 50 kDa and cleaved to the mature form of the

enzyme by a metalloprotease, although the precursor form is also active. The copper atom in LO is involved in the transport of electrons to and from oxygen to facilitate the oxidative deamination of lysine residues in these extracellular matrix proteins. While the coordination of copper is essential to LO activity, insufficient dietary intake of copper does not influence the expression of the apoenzyme. However, the absence of the functional LO is linked to the skeletal and vascular tissue disorders that are associated with dietary copper deficiency. LO is also inhibited by a variety of semicarbazides, hydrazines, and amino nitrites, as well as heparin. Beta-aminopropionitrile is a commonly used inhibitor. LO activity is increased in response to ozone, cadmium, and elevated levels of hormones released in response to local tissue trauma, such as transforming growth factor-beta, platelet-derived growth factor, angiotensin II, and fibroblast growth factor. Abnormalities in LO activity have been linked to Menkes syndrome and occipital horn syndrome. Cytosolic forms of the enzyme have been implicated in abnormal cell proliferation (reviewed in Rucker, R.B. et al. (1998) Am. J. Clin. Nutr. 67:996S-1002S and Smith-Mungo, L.I. and Kagan, H.M. (1998) Matrix Biol. 16:387-398).

Dihydrofolate reductases

Dihydrofolate reductases (DHFR) are ubiquitous enzymes that catalyze the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, an essential step in the de novo synthesis of glycine and purines as well as the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). The basic reaction is as follows:

$$\text{7,8-dihydrofolate} + \text{NADPH} \rightarrow \text{5,6,7,8-tetrahydrofolate} + \text{NADP}^+$$

The enzymes can be inhibited by a number of dihydrofolate analogs, including trimethroprim and methotrexate. Since an abundance of dTMP is required for DNA synthesis, rapidly dividing cells require the activity of DHFR. The replication of DNA viruses (i.e., herpesvirus) also requires high levels of DHFR activity. As a result, drugs that target DHFR have been used for cancer chemotherapy and to inhibit DNA virus replication. (For similar reasons, thymidylate synthetases are also target enzymes.) Drugs that inhibit DHFR are preferentially cytotoxic for rapidly dividing cells (or DNA virus-infected cells) but have no specificity, resulting in the indiscriminate destruction of dividing cells. Furthermore, cancer cells may become resistant to drugs such as methotrexate as a result of acquired transport defects or the duplication of one or more DHFR genes (Stryer, L. (1988) Biochemistry. W.H Freeman and Co., Inc. New York. pp. 511-5619).

Aldo/keto reductases

Aldo/keto reductases are monomeric NADPH-dependent oxidoreductases with broad substrate specificities (Bohren, K.M. et al. (1989) J. Biol. Chem. 264:9547-9551). These enzymes catalyze the reduction of carbonyl-containing compounds, including carbonyl-containing sugars and

aromatic compounds, to the corresponding alcohols. Therefore, a variety of carbonyl-containing drugs and xenobiotics are likely metabolized by enzymes of this class.

One known reaction catalyzed by a family member, aldose reductase, is the reduction of glucose to sorbitol, which is then further metabolized to fructose by sorbitol dehydrogenase. Under normal conditions, the reduction of glucose to sorbitol is a minor pathway. In hyperglycemic states, however, the accumulation of sorbitol is implicated in the development of diabetic complications (OMIM *103880 Aldo-keto reductase family 1, member B1). Members of this enzyme family are also highly expressed in some liver cancers (Cao, D. et al. (1998) J. Biol. Chem. 273:11429-11435).

Another important member of the aldo-keto reductase family is $NADP^+$-dependent morphine 6-dehydrogenase (MDH). MDH catalyses the oxidation of the 6-hydroxyl groups of morphine and codeine to ketones, resulting in the formation of morphinone and codeinone. Oxidation by MDH represents one of several pathways for the metabolism of the opiates. Other pathways involve glucuronidation, *O*-demethylation and N-dealkylation. Active opiate metabolites are produced by several of these enzymatic reactions (Cone, E.J. et al. (1978) Drug. Metab. Dispos. 6:488-93 and Tegeder, I. et al. (1999) Clin. Pharmacokinet. 37:17-40). Purified MDH has shown promise in detecting the presence of opiates in biological samples (Holt, P.J. et al. (1996) Anal. Chem. 68:1877-82).

Alcohol dehydrogenases

Alcohol dehydrogenases (ADHs) oxidize simple alcohols to the corresponding aldehydes. ADH is a cytosolic enzyme, prefers the cofactor NAD^+, and also binds zinc ion. Liver contains the highest levels of ADH, with lower levels in kidney, lung, and the gastric mucosa.

Known ADH isoforms are dimeric proteins composed of 40 kDa subunits. There are five known gene loci which encode these subunits (a, b, g, p, c), and some of the loci have characterized allelic variants (b_1, b_2, b_3, g_1, g_2). The subunits can form homodimers and heterodimers; the subunit composition determines the specific properties of the active enzyme. The holoenzymes have therefore been categorized as Class I (subunit compositions aa, ab, ag, bg, gg), Class II (pp), and Class III (cc). Class I ADH isozymes oxidize ethanol and other small aliphatic alcohols, and are inhibited by pyrazole. Class II isozymes prefer longer chain aliphatic and aromatic alcohols, are unable to oxidize methanol, and are not inhibited by pyrazole. Class III isozymes prefer even longer chain aliphatic alcohols (five carbons and longer) and aromatic alcohols, and are not inhibited by pyrazole.

The short-chain alcohol dehydrogenases include a number of related enzymes with a variety of substrate specificities. Included in this group are the mammalian enzymes D-beta-hydroxybutyrate dehydrogenase, (R)-3-hydroxybutyrate dehydrogenase, 15-hydroxyprostaglandin dehydrogenase, NADPH-dependent carbonyl reductase, corticosteroid 11-beta-dehydrogenase, and estradiol 17-beta-

dehydrogenase, as well as the bacterial enzymes acetoacetyl-CoA reductase, glucose 1-dehydrogenase, 3-beta-hydroxysteroid dehydrogenase, 20-beta-hydroxysteroid dehydrogenase, ribitol dehydrogenase, 3-oxoacyl reductase, 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase, sorbitol-6-phosphate 2-dehydrogenase, 7-alpha-hydroxysteroid dehydrogenase, cis-1,2-dihydroxy-3,4-cyclohexadiene-1-carboxylate dehydrogenase, *cis*-toluene dihydrodiol dehydrogenase, *cis*-benzene glycol dehydrogenase, biphenyl-2,3-dihydro-2,3-diol dehydrogenase, N-acylmannosamine 1-dehydrogenase, and 2-deoxy-D-gluconate 3-dehydrogenase (Krozowski, Z. (1994) J. Steroid Biochem. Mol. Biol. 51:125-130; Krozowski, Z. (1992) Mol. Cell Endocrinol. 84:C25-31; and Marks, A.R. et al. (1992) J. Biol. Chem. 267:15459-15463).

UDP glucuronyltransferase

Members of the UDP glucuronyltransferase family (UGTs) catalyze the transfer of a glucuronic acid group from the cofactor uridine diphosphate-glucuronic acid (UDP-glucuronic acid) to a substrate. The transfer is generally to a nucleophilic heteroatom (O, N, or S). Substrates include xenobiotics which have been functionalized by Phase I reactions, as well as endogenous compounds such as bilirubin, steroid hormones, and thyroid hormones. Products of glucuronidation are excreted in urine if the molecular weight of the substrate is less than about 250 g/mol, whereas larger glucuronidated substrates are excreted in bile.

UGTs are located in the microsomes of liver, kidney, intestine, skin, brain, spleen, and nasal mucosa, where they are on the same side of the endoplasmic reticulum membrane as cytochrome P450 enzymes and flavin-containing monooxygenases, and therefore are ideally located to access products of Phase I drug metabolism. UGTs have a C-terminal membrane-spanning domain which anchors them in the endoplasmic reticulum membrane, and a conserved signature domain of about 50 amino acid residues in their C terminal section (Prosite PDOC00359 UDP-glycosyltransferase signature).

UGTs involved in drug metabolism are encoded by two gene families, UGT1 and UGT2. Members of the UGT1 family result from alternative splicing of a single gene locus, which has a variable substrate binding domain and constant region involved in cofactor binding and membrane insertion. Members of the UGT2 family are encoded by separate gene loci, and are divided into two families, UGT2A and UGT2B. The 2A subfamily is expressed in olfactory epithelium, and the 2B subfamily is expressed in liver microsomes. Mutations in UGT genes are associated with hyperbilirubinemia (OMIM #143500 Hyperbilirubinemia I); Crigler-Najjar syndrome, characterized by intense hyperbilirubinemia from birth (OMIM #218800 Crigler-Najjar syndrome); and a milder form of hyperbilirubinemia termed Gilbert's disease (OMIM *191740 UGT1).

Sulfotransferase

Sulfate conjugation occurs on many of the same substrates which undergo O-glucuronidation

to produce a highly water-soluble sulfuric acid ester. Sulfotransferases (ST) catalyze this reaction by transferring SO_3^- from the cofactor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to the substrate. ST substrates are predominantly phenols and aliphatic alcohols, but also include aromatic amines and aliphatic amines, which are conjugated to produce the corresponding sulfamates. The products of these reactions are excreted mainly in urine.

STs are found in a wide range of tissues, including liver, kidney, intestinal tract, lung, platelets, and brain. The enzymes are generally cytosolic, and multiple forms are often co-expressed. For example, there are more than a dozen forms of ST in rat liver cytosol. These biochemically characterized STs fall into five classes based on their substrate preference: arylsulfotransferase, alcohol sulfotransferase, estrogen sulfotransferase, tyrosine ester sulfotransferase, and bile salt sulfotransferase.

ST enzyme activity varies greatly with sex and age in rats. The combined effects of developmental cues and sex-related hormones are thought to lead to these differences in ST expression profiles, as well as the profiles of other DMEs such as cytochromes P450. Notably, the high expression of STs in cats partially compensates for their low level of UDP glucuronyltransferase activity.

Several forms of ST have been purified from human liver cytosol and cloned. There are two phenol sulfotransferases with different thermal stabilities and substrate preferences. The thermostable enzyme catalyzes the sulfation of phenols such as para-nitrophenol, minoxidil, and acetaminophen; the thermolabile enzyme prefers monoamine substrates such as dopamine, epinephrine, and levadopa. Other cloned STs include an estrogen sulfotransferase and an N-acetylglucosamine-6-O-sulfotransferase. This last enzyme is illustrative of the other major role of STs in cellular biochemistry, the modification of carbohydrate structures that may be important in cellular differentiation and maturation of proteoglycans. Indeed, an inherited defect in a sulfotransferase has been implicated in macular corneal dystrophy, a disorder characterized by a failure to synthesize mature keratan sulfate proteoglycans (Nakazawa, K. et al. (1984) J. Biol. Chem. 259:13751-13757; OMIM *217800 Macular dystrophy, corneal).

Galactosyltransferases

Galactosyltransferases are a subset of glycosyltransferases that transfer galactose (Gal) to the terminal N-acetylglucosamine (GlcNAc) oligosaccharide chains that are part of glycoproteins or glycolipids that are free in solution (Kolbinger, F. et al. (1998) J. Biol. Chem. 273:433-440; Amado, M. et al. (1999) Biochim. Biophys. Acta 1473:35-53). Galactosyltransferases have been detected on the cell surface and as soluble extracellular proteins, in addition to being present in the Golgi. β1,3-galactosyltransferases form Type I carbohydrate chains with Gal (β1-3)GlcNAc linkages. Known human and mouse β1,3-galactosyltransferases appear to have a short cytosolic domain, a single

transmembrane domain, and a catalytic domain with eight conserved regions. (Kolbinger, supra and Hennet, T. et al. (1998) J. Biol. Chem. 273:58-65). In mouse UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 1 is located at amino acid residues 78-83, region 2 is located at amino acid residues 93-102, region 3 is located at amino acid residues 116-119, region 4 is located at amino acid residues 147-158, region 5 is located at amino acid residues 172-183, region 6 is located at amino acid residues 203-206, region 7 is located at amino acid residues 236-246, and region 8 is located at amino acid residues 264-275. A variant of a sequence found within mouse UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 8 is also found in bacterial galactosyltransferases, suggesting that this sequence defines a galactosyltransferase sequence motif (Hennet, supra). Recent work suggests that brainiac protein is a β1,3-galactosyltransferase (Yuan, Y. et al. (1997) Cell 88:9-11; and Hennet, supra).

UDP-Gal:GlcNAc-1,4-galactosyltransferase (-1,4-GalT) (Sato, T. et al., (1997) EMBO J. 16:1850-1857) catalyzes the formation of Type II carbohydrate chains with Gal (β1-4)GlcNAc linkages. As is the case with the β1,3-galactosyltransferase, a soluble form of the enzyme is formed by cleavage of the membrane-bound form. Amino acids conserved among β1,4-galactosyltransferases include two cysteines linked through a disulfide-bond and a putative UDP-galactose-binding site in the catalytic domain (Yadav, S. and Brew, K. (1990) J. Biol. Chem. 265:14163-14169; Yadav, S.P. and Brew, K. (1991) J. Biol. Chem. 266:698-703; and Shaper, N.L. et al. (1997) J. Biol. Chem. 272:31389-31399). β1,4-galactosyltransferases have several specialized roles in addition to synthesizing carbohydrate chains on glycoproteins or glycolipids. In mammals a β1,4-galactosyltransferase, as part of a heterodimer with α-lactalbumin, functions in lactating mammary gland lactose production. A β1,4-galactosyltransferase on the surface of sperm functions as a receptor that specifically recognizes the egg. Cell surface β1,4-galactosyltransferases also function in cell adhesion, cell/basal lamina interaction, and normal and metastatic cell migration. (Shur, B. (1993) Curr. Opin. Cell Biol. 5:854-863; and Shaper, J. (1995) Adv. Exp. Med. Biol. 376:95-104).

Glutathione S-transferase

The basic reaction catalyzed by glutathione S-transferases (GST) is the conjugation of an electrophile with reduced glutathione (GSH). GSTs are homodimeric or heterodimeric proteins localized mainly in the cytosol, but some level of activity is present in microsomes as well. The major isozymes share common structural and catalytic properties; in humans they have been classified into four major classes, Alpha, Mu, Pi, and Theta. The two largest classes, Alpha and Mu, are identified by their respective protein isoelectric points; pI ~ 7.5-9.0 (Alpha), and pI ~ 6.6 (Mu). Each GST possesses a common binding site for GSH and a variable hydrophobic binding site. The hydrophobic binding site in each isozyme is specific for particular electrophilic substrates. Specific amino acid residues within GSTs have been identified as important for these binding sites and for

catalytic activity. Residues Q67, T68, D101, E104, and R131 are important for the binding of GSH (Lee, H.-C. et al. (1995) J. Biol. Chem. 270:99-109). Residues R13, R20, and R69 are important for the catalytic activity of GST (Stenberg, G. et al. (1991) Biochem. J. 274:549-555).

In most cases, GSTs perform the beneficial function of deactivation and detoxification of potentially mutagenic and carcinogenic chemicals. However, in some cases their action is detrimental and results in activation of chemicals with consequent mutagenic and carcinogenic effects. Some forms of rat and human GSTs are reliable preneoplastic markers that aid in the detection of carcinogenesis. Expression of human GSTs in bacterial strains, such as Salmonella typhimurium used in the well-known Ames test for mutagenicity, has helped to establish the role of these enzymes in mutagenesis. Dihalomethanes, which produce liver tumors in mice, are believed to be activated by GST. This view is supported by the finding that dihalomethanes are more mutagenic in bacterial cells expressing human GST than in untransfected cells (Thier, R. et al. (1993) Proc. Natl. Acad. Sci. USA 90:8567-8580). The mutagenicity of ethylene dibromide and ethylene dichloride is increased in bacterial cells expressing the human Alpha GST, A1-1, while the mutagenicity of aflatoxin B1 is substantially reduced by enhancing the expression of GST (Simula, T.P. et al. (1993) Carcinogenesis 14:1371-1376). Thus, control of GST activity may be useful in the control of mutagenesis and carcinogenesis.

GST has been implicated in the acquired resistance of many cancers to drug treatment, the phenomenon known as multi-drug resistance (MDR). MDR occurs when a cancer patient is treated with a cytotoxic drug such as cyclophosphamide and subsequently becomes resistant to this drug and to a variety of other cytotoxic agents as well. Increased GST levels are associated with some of these drug resistant cancers, and it is believed that this increase occurs in response to the drug agent which is then deactivated by the GST catalyzed GSH conjugation reaction. The increased GST levels then protect the cancer cells from other cytotoxic agents which bind to GST. Increased levels of A1-1 in tumors has been linked to drug resistance induced by cyclophosphamide treatment (Dirven H.A. et al. (1994) Cancer Res. 54: 6215-6220). Thus control of GST activity in cancerous tissues may be useful in treating MDR in cancer patients.

Gamma-glutamyl transpeptidase

Gamma-glutamyl transpeptidases are ubiquitously expressed enzymes that initiate extracellular glutathione (GSH) breakdown by cleaving gamma-glutamyl amide bonds. The breakdown of GSH provides cells with a regional cysteine pool for biosynthetic pathways. Gamma-glutamyl transpeptidases also contribute to cellular antioxidant defenses and expression is induced by oxidative stress. The cell surface-localized glycoproteins are expressed at high levels in cancer cells. Studies have suggested that the high level of gamma-glutamyl transpeptidase activity present on the surface of cancer cells could be exploited to activate precursor drugs, resulting in high

local concentrations of anti-cancer therapeutic agents (Hanigan, M.H. (1998) Chem. Biol. Interact. 111-112:333-42; Taniguchi, N. and Ikeda, Y. (1998) Adv. Enzymol. Relat. Areas Mol. Biol. 72:239-78; Chikhi, N. et al. (1999) Comp. Biochem. Physiol. B. Biochem. Mol. Biol. 122:367-380).

Acyltransferase

N-acyltransferase enzymes catalyze the transfer of an amino acid conjugate to an activated carboxylic group. Endogenous compounds and xenobiotics are activated by acyl-CoA synthetases in the cytosol, microsomes, and mitochondria. The acyl-CoA intermediates are then conjugated with an amino acid (typically glycine, glutamine, or taurine, but also ornithine, arginine, histidine, serine, aspartic acid, and several dipeptides) by N-acyltransferases in the cytosol or mitochondria to form a metabolite with an amide bond. This reaction is complementary to O-glucuronidation, but amino acid conjugation does not produce the reactive and toxic metabolites which often result from glucuronidation.

One well-characterized enzyme of this class is the bile acid-CoA:amino acid N-acyltransferase (BAT) responsible for generating the bile acid conjugates which serve as detergents in the gastrointestinal tract (Falany, C.N. et al. (1994) J. Biol. Chem. 269:19375-19379; Johnson, M.R. et al. (1991) J. Biol. Chem. 266:10227-10233). BAT is also useful as a predictive indicator for prognosis of hepatocellular carcinoma patients after partial hepatectomy (Furutani, M. et al. (1996) Hepatology 24:1441-1445).

Acetyltransferases

Acetyltransferases have been extensively studied for their role in histone acetylation. Histone acetylation results in the relaxing of the chromatin structure in eukaryotic cells, allowing transcription factors to gain access to promoter elements of the DNA templates in the affected region of the genome (or the genome in general). In contrast, histone deacetylation results in a reduction in transcription by closing the chromatin structure and limiting access of transcription factors. To this end, a common means of stimulating cell transcription is the use of chemical agents that inhibit the deacetylation of histones (e.g., sodium butyrate), resulting in a global (albeit artifactual) increase in gene expression. The modulation of gene expression by acetylation also results from the acetylation of other proteins, including but not limited to, p53, GATA-1, MyoD, ACTR, TFIIE, TFIIF and the high mobility group proteins (HMG). In the case of p53, acetylation results in increased DNA binding, leading to the stimulation of transcription of genes regulated by p53. The prototypic histone acetylase (HAT) is Gcn5 from Saccharomyces cerevisiae. Gcn5 is a member of a family of acetylases that includes Tetrahymena p55, human Gcn5, and human p300/CBP. Histone acetylation is reviewed in (Cheung, W.L. et al. (2000) Curr. Opin. Cell Biol. 12:326-333 and Berger, S.L (1999) Curr. Opin. Cell Biol. 11:336-341). Some acetyltransferase enzymes possess the alpha/beta hydrolase fold (Center of Applied Molecular Engineering Inst. of Chemistry and Biochemistry - University of

Salzburg, http://predict.sanger.ac.uk/irbm-course97/Docs/ms/) common to several other major classes of enzymes, including but not limited to, acetylcholinesterases and carboxylesterases (Structural Classification of Proteins, http://scop.mrc-lmb.cam.ac.uk/scop/index.html).

N-acetyltransferase

Aromatic amines and hydrazine-containing compounds are subject to N-acetylation by the N-acetyltransferase enzymes of liver and other tissues. Some xenobiotics can be O-acetylated to some extent by the same enzymes. N-acetyltransferases are cytosolic enzymes which utilize the cofactor acetyl-coenzyme A (acetyl-CoA) to transfer the acetyl group in a two step process. In the first step, the acetyl group is transferred from acetyl-CoA to an active site cysteine residue; in the second step, the acetyl group is transferred to the substrate amino group and the enzyme is regenerated.

In contrast to most other DME classes, there are a limited number of known N-acetyltransferases. In humans, there are two highly similar enzymes, NAT1 and NAT2; mice appear to have a third form of the enzyme, NAT3. The human forms of N-acetyltransferase have independent regulation (NAT1 is widely-expressed, whereas NAT2 is in liver and gut only) and overlapping substrate preferences. Both enzymes appear to accept most substrates to some extent, but NAT1 does prefer some substrates (para-aminobenzoic acid, para-aminosalicylic acid, sulfamethoxazole, and sulfanilamide), while NAT2 prefers others (isoniazid, hydralazine, procainamide, dapsone, aminoglutethimide, and sulfamethazine).

Clinical observations of patients taking the antituberculosis drug isoniazid in the 1950s led to the description of fast and slow acetylators of the compound. These phenotypes were shown subsequently to be due to mutations in the NAT2 gene which affected enzyme activity or stability. The slow isoniazid acetylator phenotype is very prevalent in Middle Eastern populations (approx. 70%), and is less prevalent in Caucasian (approx. 50%) and Asian (<25%) populations. More recently, functional polymorphism in NAT1 has been detected, with approximately 8% of the population tested showing a slow acetylator phenotype (Butcher, N. J. et al. (1998) Pharmacogenetics 8:67-72). Since NAT1 can activate some known aromatic amine carcinogens, polymorphism in the widely-expressed NAT1 enzyme may be important in determining cancer risk (OMIM *108345 N-acetyltransferase 1).

Aminotransferases

Aminotransferases comprise a family of pyridoxal 5'-phosphate (PLP) -dependent enzymes that catalyze transformations of amino acids. Aspartate aminotransferase (AspAT) is the most extensively studied PLP-containing enzyme. It catalyzes the reversible transamination of dicarboxylic L-amino acids, aspartate and glutamate, and the corresponding 2-oxo acids, oxalacetate and 2-oxoglutarate. Other members of the family include pyruvate aminotransferase, branched-chain amino acid aminotransferase, tyrosine aminotransferase, aromatic aminotransferase,

alanine:glyoxylate aminotransferase (AGT), and kynurenine aminotransferase (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937).

Primary hyperoxaluria type-1 is an autosomal recessive disorder resulting in a deficiency in the liver-specific peroxisomal enzyme, alanine:glyoxylate aminotransferase-1. The phenotype of the disorder is a deficiency in glyoxylate metabolism. In the absence of AGT, glyoxylate is oxidized to oxalate rather than being transaminated to glycine. The result is the deposition of insoluble calcium oxalate in the kidneys and urinary tract, ultimately causing renal failure (Lumb, M.J. et al. (1999) J. Biol. Chem. 274:20587-20596).

Kynurenine aminotransferase catalyzes the irreversible transamination of the L-tryptophan metabolite L-kynurenine to form kynurenic acid. The enzyme may also catalyze the reversible transamination reaction between L-2-aminoadipate and 2-oxoglutarate to produce 2-oxoadipate and L-glutamate. Kynurenic acid is a putative modulator of glutamatergic neurotransmission; thus a deficiency in kynurenine aminotransferase may be associated with pleotrophic effects (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

Catechol-*O*-methyltransferase

Catechol-*O*-methyltransferase (COMT) catalyzes the transfer of the methyl group of *S*-adenosyl-L-methionine (AdoMet; SAM) donor to one of the hydroxyl groups of the catechol substrate (e.g., L-dopa, dopamine, or DBA). Methylation of the 3'-hydroxyl group is favored over methylation of the 4'-hydroxyl group and the membrane bound isoform of COMT is more regiospecific than the soluble form. Translation of the soluble form of the enzyme results from utilization of an internal start codon in a full-length mRNA (1.5 kb) or from the translation of a shorter mRNA (1.3 kb), transcribed from an internal promoter. The proposed S_N2-like methylation reaction requires Mg^{++} and is inhibited by Ca^{++}. The binding of the donor and substrate to COMT occurs sequentially. AdoMet first binds COMT in a Mg^{++}-independent manner, followed by the binding of Mg^{++} and the binding of the catechol substrate.

The amount of COMT in tissues is relatively high compared to the amount of activity normally required, thus inhibition is problematic. Nonetheless, inhibitors have been developed for in vitro use (e.g., gallates, tropolone, U-0521, and 3',4'-dihydroxy-2-methyl-propiophetropolone) and for clinical use (e.g., nitrocatechol-based compounds and tolcapone). Administration of these inhibitors results in the increased half-life of L-dopa and the consequent formation of dopamine. Inhibition of COMT is also likely to increase the half-life of various other catechol-structure compounds, including but not limited to epinephrine/norepinephrine, isoprenaline, rimiterol, dobutamine, fenoldopam, apomorphine, and α-methyldopa. A deficiency in norepinephrine has been linked to clinical depression, hence the use of COMT inhibitors could be usefull in the treatment of depression. COMT inhibitors are generally well tolerated with minimal side effects and are ultimately

metabolized in the liver with only minor accumulation of metabolites in the body (Männistö, P.T. and Kaakkola, S. (1999) Pharmacol. Rev. 51:593-628).

Copper-zinc superoxide dismutases

Copper-zinc superoxide dismutases are compact homodimeric metalloenzymes involved in cellular defenses against oxidative damage. The enzymes contain one atom of zinc and one atom of copper per subunit and catalyze the dismutation of superoxide anions into O_2 and H_2O_2. The rate of dismutation is diffusion-limited and consequently enhanced by the presence of favorable electrostatic interactions between the substrate and enzyme active site. Examples of this class of enzyme have been identified in the cytoplasm of all the eukaryotic cells as well as in the periplasm of several bacterial species. Copper-zinc superoxide dismutases are robust enzymes that are highly resistant to proteolytic digestion and denaturing by urea and SDS. In addition to the compact structure of the enzymes, the presence of the metal ions and intrasubunit disulfide bonds is believed to be responsible for enzyme stability. The enzymes undergo reversible denaturation at temperatures as high as 70°C (Battistoni, A. et al. (1998) J. Biol. Chem. 273:5655-5661).

Overexpression of superoxide dismutase has been implicated in enhancing freezing tolerance of transgenic alfalfa as well as providing resistance to environmental toxins such as the diphenyl ether herbicide, acifluorfen (McKersie, B.D. et al. (1993) Plant Physiol. 103:1155-1163). In addtion, yeast cells become more resistant to freeze-thaw damage following exposure to hydrogen peroxide which causes the yeast cells to adapt to further peroxide stress by upregulating expression of superoxide dismutases. In this study, mutations to yeast superoxide dismutase genes had a more detrimental effect on freeze-thaw resistance than mutations which affected the regulation of glutathione metabolism, long suspected of being important in determining an organism's survival through the process of cryopreservation (Jong-In Park, J.-I. et al. (1998) J. Biol. Chem. 273:22921-22928).

Expression of superoxide dismutase is also associated with Mycobacterium tuberculosis, the organism that causes tuberculosis. Superoxide dismutase is one of the ten major proteins secreted by M. tuberculosis and its expression is upregulated approximately 5-fold in response to oxidative stress. M. tuberculosis expresses almost two orders of magnitude more superoxide dismutase than the nonpathogenic mycobacterium M. smegmatis, and secretes a much higher proportion of the expressed enzyme. The result is the secretion of ~350-fold more enzyme by M. tuberculosis than M. smegmatis, providing substantial resistance to oxidative stress (Harth, G. and Horwitz, M.A. (1999) J. Biol. Chem. 274:4281-4292).

The reduced expression of copper-zinc superoxide dismutases, as well as other enzymes with anti-oxidant capabilities, has been implicated in the early stages of cancer. The expression of copper-zinc superoxide dismutases has been shown to be lower in prostatic intraepithelial neoplasia and prostate carcinomas, compared to normal prostate tissue (Bostwick, D.G. (2000) Cancer 89:123-134).

Phosphodiesterases

Phosphodiesterases make up a class of enzymes which catalyze the hydrolysis of one of the two ester bonds in a phosphodiester compound. Phosphodiesterases are therefore crucial to a variety of cellular processes. Phosphodiesterases include DNA and RNA endonucleases and exonucleases, which are essential for cell growth and replication, and topoisomerases, which break and rejoin nucleic acid strands during topological rearrangement of DNA. A Tyr-DNA phosphodiesterase functions in DNA repair by hydrolyzing dead-end covalent intermediates formed between topoisomerase I and DNA (Pouliot, J.J. et al. (1999) Science 286:552-555; Yang, S.-W. (1996) Proc. Natl. Acad. Sci. USA 93:11534-11539).

Acid sphingomyelinase is a phosphodiesterase which hydrolyzes the membrane phospholipid sphingomyelin to produce ceramide and phosphorylcholine. Phosphorylcholine is used in the synthesis of phosphatidylcholine, which is involved in numerous intracellular signaling pathways, while ceramide is an essential precursor for the generation of gangliosides, membrane lipids found in high concentration in neural tissue. Defective acid sphingomyelinase leads to a build-up of sphingomyelin molecules in lysosomes, resulting in Niemann-Pick disease (Schuchman, E.H. and S.R. Miranda (1997) Genet. Test. 1:13-19).

Glycerophosphoryl diester phosphodiesterase (also known as glycerophosphodiester phosphodiesterase) is a phosphodiesterase which hydrolyzes deacetylated phospholipid glycerophosphodiesters to produce sn-glycerol-3-phosphate and an alcohol. Glycerophosphocholine, glycerophosphoethanolamine, glycerophosphoglycerol, and glycerophosphoinositol are examples of substrates for glycerophosphoryl diester phosphodiesterases. A glycerophosphoryl diester phosphodiesterase from E. coli has broad specificity for glycerophosphodiester substrates (Larson, T.J. et al. (1983) J. Biol. Chem. 248:5428-5432).

Cyclic nucleotide phosphodiesterases (PDEs) are crucial enzymes in the regulation of the cyclic nucleotides cAMP and cGMP. cAMP and cGMP function as intracellular second messengers to transduce a variety of extracellular signals including hormones, light, and neurotransmitters. PDEs degrade cyclic nucleotides to their corresponding monophosphates, thereby regulating the intracellular concentrations of cyclic nucleotides and their effects on signal transduction. Due to their roles as regulators of signal transduction, PDEs have been extensively studied as chemotherapeutic targets (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481; Torphy, J.T. (1998) Am. J. Resp. Crit. Care Med. 157:351-370).

Families of mammalian PDEs have been classified based on their substrate specificity and affinity, sensitivity to cofactors, and sensitivity to inhibitory agents (Beavo, J.A. (1995) Physiol. Rev. 75:725-748; Conti, M. et al. (1995) Endocrine Rev. 16:370-389). Several of these families contain distinct genes, many of which are expressed in different tissues as splice variants. Within PDE

families, there are multiple isozymes and multiple splice variants of these isozymes (Conti, M. and S.-L.C. Jin (1999) Prog. Nucleic Acid Res. Mol. Biol. 63:1-38). The existence of multiple PDE families, isozymes, and splice variants is an indication of the variety and complexity of the regulatory pathways involving cyclic nucleotides (Houslay, M.D. and G. Milligan (1997) Trends Biochem. Sci. 22:217-224).

Type 1 PDEs (PDE1s) are Ca^{2+}/calmodulin-dependent and appear to be encoded by at least three different genes, each having at least two different splice variants (Kakkar, R. et al. (1999) Cell Mol. Life Sci. 55:1164-1186). PDE1s have been found in the lung, heart, and brain. Some PDE1 isozymes are regulated in vitro by phosphorylation/dephosphorylation. Phosphorylation of these PDE1 isozymes decreases the affinity of the enzyme for calmodulin, decreases PDE activity, and increases steady state levels of cAMP (Kakkar, supra). PDE1s may provide useful therapeutic targets for disorders of the central nervous system and the cardiovascular and immune systems, due to the involvement of PDE1s in both cyclic nucleotide and calcium signaling (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481).

PDE2s are cGMP-stimulated PDEs that have been found in the cerebellum, neocortex, heart, kidney, lung, pulmonary artery, and skeletal muscle (Sadhu, K. et al. (1999) J. Histochem. Cytochem. 47:895-906). PDE2s are thought to mediate the effects of cAMP on catecholamine secretion, participate in the regulation of aldosterone (Beavo, supra), and play a role in olfactory signal transduction (Juilfs, D.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:3388-3395).

PDE3s have high affinity for both cGMP and cAMP, and so these cyclic nucleotides act as competitive substrates for PDE3s. PDE3s play roles in stimulating myocardial contractility, inhibiting platelet aggregation, relaxing vascular and airway smooth muscle, inhibiting proliferation of T-lymphocytes and cultured vascular smooth muscle cells, and regulating catecholamine-induced release of free fatty acids from adipose tissue. The PDE3 family of phosphodiesterases are sensitive to specific inhibitors such as cilostamide, enoximone, and lixazinone. Isozymes of PDE3 can be regulated by cAMP-dependent protein kinase, or by insulin-dependent kinases (Degerman, E. et al. (1997) J. Biol. Chem. 272:6823-6826).

PDE4s are specific for cAMP; are localized to airway smooth muscle, the vascular endothelium, and all inflammatory cells; and can be activated by cAMP-dependent phosphorylation. Since elevation of cAMP levels can lead to suppression of inflammatory cell activation and to relaxation of bronchial smooth muscle, PDE4s have been studied extensively as possible targets for novel anti-inflammatory agents, with special emphasis placed on the discovery of asthma treatments. PDE4 inhibitors are currently undergoing clinical trials as treatments for asthma, chronic obstructive pulmonary disease, and atopic eczema. All four known isozymes of PDE4 are susceptible to the inhibitor rolipram, a compound which has been shown to improve behavioral memory in mice (Barad,

M. et al. (1998) Proc. Natl. Acad. Sci. USA 95:15020-15025). PDE4 inhibitors have also been studied as possible therapeutic agents against acute lung injury, endotoxemia, rheumatoid arthritis, multiple sclerosis, and various neurological and gastrointestinal indications (Doherty, A.M. (1999) Curr. Opin. Chem. Biol. 3:466-473).

PDE5 is highly selective for cGMP as a substrate (Turko, I.V. et al. (1998) Biochemistry 37:4200-4205), and has two allosteric cGMP-specific binding sites (McAllister-Lucas, L.M. et al. (1995) J. Biol. Chem. 270:30671-30679). Binding of cGMP to these allosteric binding sites seems to be important for phosphorylation of PDE5 by cGMP-dependent protein kinase rather than for direct regulation of catalytic activity. High levels of PDE5 are found in vascular smooth muscle, platelets, lung, and kidney. The inhibitor zaprinast is effective against PDE5 and PDE1s. Modification of zaprinast to provide specificity against PDE5 has resulted in sildenafil (VIAGRA; Pfizer, Inc., New York NY), a treatment for male erectile dysfunction (Terrett, N. et al. (1996) Bioorg. Med. Chem. Lett. 6:1819-1824). Inhibitors of PDE5 are currently being studied as agents for cardiovascular therapy (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481).

PDE6s, the photoreceptor cyclic nucleotide phosphodiesterases, are crucial components of the phototransduction cascade. In association with the G-protein transducin, PDE6s hydrolyze cGMP to regulate cGMP-gated cation channels in photoreceptor membranes. In addition to the cGMP-binding active site, PDE6s also have two high-affinity cGMP-binding sites which are thought to play a regulatory role in PDE6 function (Artemyev, N.O. et al. (1998) Methods 14:93-104). Defects in PDE6s have been associated with retinal disease. Retinal degeneration in the rd mouse (Yan, W. et al. (1998) Invest. Opthalmol. Vis. Sci. 39:2529-2536), autosomal recessive retinitis pigmentosa in humans (Danciger, M. et al. (1995) Genomics 30:1-7), and rod/cone dysplasia 1 in Irish Setter dogs (Suber, M.L. et al. (1993) Proc. Natl. Acad. Sci. USA 90:3968-3972) have been attributed to mutations in the PDE6B gene.

The PDE7 family of PDEs consists of only one known member having multiple splice variants (Bloom, T.J. and J.A. Beavo (1996) Proc. Natl. Acad. Sci. USA 93:14188-14192). PDE7s are cAMP specific, but little else is known about their physiological function. Although mRNAs encoding PDE7s are found in skeletal muscle, heart, brain, lung, kidney, and pancreas, expression of PDE7 proteins is restricted to specific tissue types (Han, P. et al. (1997) J. Biol. Chem. 272:16152-16157; Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481). PDE7s are very closely related to the PDE4 family; however, PDE7s are not inhibited by rolipram, a specific inhibitor of PDE4s (Beavo, supra).

PDE8s are cAMP specific, and are closely related to the PDE4 family. PDE8s are expressed in thyroid gland, testis, eye, liver, skeletal muscle, heart, kidney, ovary, and brain. The cAMP-hydrolyzing activity of PDE8s is not inhibited by the PDE inhibitors rolipram, vinpocetine, milrinone,

IBMX (3-isobutyl-1-methylxanthine), or zaprinast, but PDE8s are inhibited by dipyridamole (Fisher, D.A. et al. (1998) Biochem. Biophys. Res. Commun. 246:570-577; Hayashi, M. et al. (1998) Biochem. Biophys. Res. Commun. 250:751-756; Soderling, S.H. et al. (1998) Proc. Natl. Acad. Sci. USA 95:8991-8996).

PDE9s are cGMP specific and most closely resemble the PDE8 family of PDEs. PDE9s are expressed in kidney, liver, lung, brain, spleen, and small intestine. PDE9s are not inhibited by sildenafil (VIAGRA; Pfizer, Inc., New York NY), rolipram, vinpocetine, dipyridamole, or IBMX (3-isobutyl-1-methylxanthine), but they are sensitive to the PDE5 inhibitor zaprinast (Fisher, D.A. et al. (1998) J. Biol. Chem. 273:15559-15564; Soderling, S.H. et al. (1998) J. Biol. Chem. 273:15553-15558).

PDE10s are dual-substrate PDEs, hydrolyzing both cAMP and cGMP. PDE10s are expressed in brain, thyroid, and testis. (Soderling, S.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:7071-7076; Fujishige, K. et al. (1999) J. Biol. Chem. 274:18438-18445; Loughney, K. et al (1999) Gene 234:109-117).

PDEs are composed of a catalytic domain of about 270-300 amino acids, an N-terminal regulatory domain responsible for binding cofactors, and, in some cases, a hydrophilic C-terminal domain of unknown function (Conti, M. and S.-L.C. Jin (1999) Prog. Nucleic Acid Res. Mol. Biol. 63:1-38). A conserved, putative zinc-binding motif has been identified in the catalytic domain of all PDEs. N-terminal regulatory domains include non-catalytic cGMP-binding domains in PDE2s, PDE5s, and PDE6s; calmodulin-binding domains in PDE1s; and domains containing phosphorylation sites in PDE3s and PDE4s. In PDE5, the N-terminal cGMP-binding domain spans about 380 amino acid residues and comprises tandem repeats of a conserved sequence motif (McAllister-Lucas, L.M. et al. (1993) J. Biol. Chem. 268:22863-22873). The NKXnD motif has been shown by mutagenesis to be important for cGMP binding (Turko, I.V. et al. (1996) J. Biol. Chem. 271:22240-22244). PDE families display approximately 30% amino acid identity within the catalytic domain; however, isozymes within the same family typically display about 85-95% identity in this region (e.g. PDE4A vs PDE4B). Furthermore, within a family there is extensive similarity (>60%) outside the catalytic domain; while across families, there is little or no sequence similarity outside this domain.

Many of the constituent functions of immune and inflammatory responses are inhibited by agents that increase intracellular levels of cAMP (Verghese, M.W. et al. (1995) Mol. Pharmacol. 47:1164-1171). A variety of diseases have been attributed to increased PDE activity and associated with decreased levels of cyclic nucleotides. For example, a form of diabetes insipidus in mice has been associated with increased PDE4 activity, an increase in low-K_m cAMP PDE activity has been reported in leukocytes of atopic patients, and PDE3 has been associated with cardiac disease.

Many inhibitors of PDEs have been identified and have undergone clinical evaluation (Perry,

M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481; Torphy, T.J. (1998) Am. J. Respir. Crit. Care Med. 157:351-370). PDE3 inhibitors are being developed as antithrombotic agents, antihypertensive agents, and as cardiotonic agents useful in the treatment of congestive heart failure. Rolipram, a PDE4 inhibitor, has been used in the treatment of depression, and other inhibitors of PDE4 are undergoing evaluation as anti-inflammatory agents. Rolipram has also been shown to inhibit lipopolysaccharide (LPS) induced TNF-α which has been shown to enhance HIV-1 replication in vitro. Therefore, rolipram may inhibit HIV-1 replication (Angel, J.B. et al. (1995) AIDS 9:1137-1144). Additionally, rolipram, based on its ability to suppress the production of cytokines such as TNF-α and β and interferon γ, has been shown to be effective in the treatment of encephalomyelitis. Rolipram may also be effective in treating tardive dyskinesia and was effective in treating multiple sclerosis in an experimental animal model (Sommer, N. et al. (1995) Nat. Med. 1:244-248; Sasaki, H. et al. (1995) Eur. J. Pharmacol. 282:71-76).

Theophylline is a nonspecific PDE inhibitor used in the treatment of bronchial asthma and other respiratory diseases. Theophylline is believed to act on airway smooth muscle function and in an anti-inflammatory or immunomodulatory capacity in the treatment of respiratory diseases (Banner, K.H. and C.P. Page (1995) Eur. Respir. J. 8:996-1000). Pentoxifylline is another nonspecific PDE inhibitor used in the treatment of intermittent claudication and diabetes-induced peripheral vascular disease. Pentoxifylline is also known to block TNF-α production and may inhibit HIV-1 replication (Angel et al., supra).

PDEs have been reported to affect cellular proliferation of a variety of cell types (Conti et al. (1995) Endocrine Rev. 16:370-389) and have been implicated in various cancers. Growth of prostate carcinoma cell lines DU145 and LNCaP was inhibited by delivery of cAMP derivatives and PDE inhibitors (Bang, Y.J. et al. (1994) Proc. Natl. Acad. Sci. USA 91:5330-5334). These cells also showed a permanent conversion in phenotype from epithelial to neuronal morphology. It has also been suggested that PDE inhibitors have the potential to regulate mesangial cell proliferation (Matousovic, K. et al. (1995) J. Clin. Invest. 96:401-410) and lymphocyte proliferation (Joulain, C. et al. (1995) J. Lipid Mediat. Cell Signal. 11:63-79). A cancer treatment has been described that involves intracellular delivery of PDEs to particular cellular compartments of tumors, resulting in cell death (Deonarain, M.P. and A.A. Epenetos (1994) Br. J. Cancer 70:786-794).

Phosphotriesterases

Phosphotriesterases (PTE, paraoxonases) are enzymes that hydrolyze toxic organophosphorus compounds and have been isolated from a variety of tissues. The enzymes appear to be lacking in birds and insects and abundant in mammals, explaining the reduced tolerance of birds and insects to organophosphorus compound (Vilanova, E. and Sogorb, M.A. (1999) Crit. Rev. Toxicol. 29:21-57). Phosphotriesterases play a central role in the detoxification of insecticides by mammals.

Phosphotriesterase activity varies among individuals and is lower in infants than adults. PTE knockout mice are markedly more sensitive to the organophosphate-based toxins diazoxon and chlorpyrifos oxon (Furlong, C.E., et al. (2000) Neurotoxicology 21:91-100). PTEs have attracted interest as enzymes capable of the detoxification of organophosphate-containing chemical waste and warfare reagents (e.g., parathion), in addition to pesticides and insecticides. Some studies have also implicated phosphotriesterase in atherosclerosis and diseases involving lipoprotein metabolism.

Thioesterases

Two soluble thioesterases involved in fatty acid biosynthesis have been isolated from mammalian tissues, one which is active only toward long-chain fatty-acyl thioesters and one which is active toward thioesters with a wide range of fatty-acyl chain-lengths. These thioesterases catalyze the chain-terminating step in the *de novo* biosynthesis of fatty acids. Chain termination involves the hydrolysis of the thioester bond which links the fatty acyl chain to the 4'-phosphopantetheine prosthetic group of the acyl carrier protein (ACP) subunit of the fatty acid synthase (Smith, S. (1981a) Methods Enzymol. 71:181-188; Smith, S. (1981b) Methods Enzymol. 71:188-200).

E. coli contains two soluble thioesterases, thioesterase I which is active only toward long-chain acyl thioesters, and thioesterase II (TEII) which has a broad chain-length specificity (Naggert, J. et al. (1991) J. Biol. Chem. 266:11044-11050). E. coli TEII does not exhibit sequence similarity with either of the two types of mammalian thioesterases which function as chain-terminating enzymes in *de novo* fatty acid biosynthesis. Unlike the mammalian thioesterases, E. coli TEII lacks the characteristic serine active site gly-X-ser-X-gly sequence motif and is not inactivated by the serine modifying agent diisopropyl fluorophosphate. However, modification of histidine 58 by iodoacetamide and diethylpyrocarbonate abolished TEII activity. Overexpression of TEII did not alter fatty acid content in E. coli, which suggests that it does not function as a chain-terminating enzyme in fatty acid biosynthesis (Naggert et al., supra). For that reason, Naggert et al. (supra) proposed that the physiological substrates for E. coli TEII may be coenzyme A (CoA)-fatty acid esters instead of ACP-phosphopanthetheine-fatty acid esters.

Carboxylesterases

Mammalian carboxylesterases constitute a multigene family expressed in a variety of tissues and cell types. Isozymes have significant sequence homology and are classified primarily on the basis of amino acid sequence. Acetylcholinesterase, butyrylcholinesterase, and carboxylesterase are grouped into the serine superfamily of esterases (B-esterases). Other carboxylesterases include thyroglobulin, thrombin, Factor IX, gliotactin, and plasminogen. Carboxylesterases catalyze the hydrolysis of ester- and amide- groups from molecules and are involved in detoxification of drugs, environmental toxins, and carcinogens. Substrates for carboxylesterases include short- and long-chain acyl-glycerols, acylcarnitine, carbonates, dipivefrin hydrochloride, cocaine, salicylates,

capsaicin, palmitoyl-coenzyme A, imidapril, haloperidol, pyrrolizidine alkaloids, steroids, p-nitrophenyl acetate, malathion, butanilicaine, and isocarboxazide. The enzymes often demonstrate low substrate specificity. Carboxylesterases are also important for the conversion of prodrugs to their respective free acids, which may be the active form of the drug (e.g., lovastatin, used to lower blood cholesterol) (reviewed in Satoh, T. and Hosokawa, M. (1998) Annu. Rev. Pharmacol. Toxicol.38:257-288).

Neuroligins are a class of molecules that (i) have N-terminal signal sequences, (ii) resemble cell-surface receptors, (iii) contain carboxylesterase domains, (iv) are highly expressed in the brain, and (v) bind to neurexins in a calcium-dependent manner. Despite the homology to carboxylesterases, neuroligins lack the active site serine residue, implying a role in substrate binding rather than catalysis (Ichtchenko, K. et al. (1996) J. Biol. Chem. 271:2676-2682).

Squalene epoxidase

Squalene epoxidase (squalene monooxygenase, SE) is a microsomal membrane-bound, FAD-dependent oxidoreductase that catalyzes the first oxygenation step in the sterol biosynthetic pathway of eukaryotic cells. Cholesterol is an essential structural component of cytoplasmic membranes acquired via the LDL receptor-mediated pathway or the biosynthetic pathway. In the latter case, all 27 carbon atoms in the cholesterol molecule are derived from acetyl-CoA (Stryer, L., supra). SE converts squalene to 2,3(*S*)-oxidosqualene, which is then converted to lanosterol and then cholesterol. The steps involved in cholesterol biosynthesis are summarized below (Stryer, L. (1988) Biochemistry. W.H Freeman and Co., Inc. New York. pp. 554-560 and Sakakibara, J. et al. (1995) 270:17-20): acetate (from Acetyl-CoA) → 3-hydoxy-3-methyl-glutaryl CoA → mevalonate → 5-phosphomevalonate → 5-pyrophosphomevalonate → isopentenyl pyrophosphate → dimethylallyl pyrophosphate → geranyl pyrophosphate → farnesyl pyrophosphate → squalene → squalene epoxide → lanosterol → cholesterol.

While cholesterol is essential for the viability of eukaryotic cells, inordinately high serum cholesterol levels result in the formation of atherosclerotic plaques in the arteries of higher organisms. This deposition of highly insoluble lipid material onto the walls of essential blood vessels (e.g., coronary arteries) results in decreased blood flow and potential necrosis of the tissues deprived of adequate blood flow. HMG-CoA reductase is responsible for the conversion of 3-hydroxyl-3-methyl-glutaryl CoA (HMG-CoA) to mevalonate, which represents the first committed step in cholesterol biosynthesis. HMG-CoA is the target of a number of pharmaceutical compounds designed to lower plasma cholesterol levels. However, inhibition of MHG-CoA also results in the reduced synthesis of non-sterol intermediates (e.g., mevalonate) required for other biochemical pathways. SE catalyzes a rate-limiting reaction that occurs later in the sterol synthesis pathway and cholesterol is the only end product of the pathway following the step catalyzed by SE. As a result, SE is the ideal target for the design of anti-hyperlipidemic drugs that do not cause a reduction in other

necessary intermediates (Nakamura, Y. et al. (1996) 271:8053-8056).

Epoxide hydrolases

Epoxide hydrolases catalyze the addition of water to epoxide-containing compounds, thereby hydrolyzing epoxides to their corresponding 1,2-diols. They are related to bacterial haloalkane dehalogenases and show sequence similarity to other members of the α/β hydrolase fold family of enzymes (e.g., bromoperoxidase A2 from Streptomyces aureofaciens, hydroxymuconic semialdehyde hydrolases from Pseudomonas putida, and haloalkane dehalogenase from Xanthobacter autotrophicus). Epoxide hydrolases are ubiquitous in nature and have been found in mammals, invertebrates, plants, fungi, and bacteria. This family of enzymes is important for the detoxification of xenobiotic epoxide compounds which are often highly electrophilic and destructive when introduced into an organism. Examples of epoxide hydrolase reactions include the hydrolysis of cis-9;10-epoxyoctadec-9(Z)-enoic acid (leukotoxin) to form its corresponding diol, threo-9,10-dihydroxyoctadec-12(Z)-enoic acid (leukotoxin diol), and the hydrolysis of cis-12,13-epoxyoctadec-9(Z)-enoic acid (isoleukotoxin) to form its corresponding diol threo-12,13-dihydroxyoctadec-9(Z)-enoic acid (isoleukotoxin diol). Leukotoxins alter membrane permeability and ion transport and cause inflammatory responses. In addition, epoxide carcinogens are known to be produced by cytochrome P450 as intermediates in the detoxification of drugs and environmental toxins.

The enzymes possess a catalytic triad composed of Asp (the nucleophile), Asp (the histidine-supporting acid), and His (the water-activating histidine). The reaction mechanism of epoxide hydrolase proceeds via a covalently bound ester intermediate initiated by the nucleophilic attack of one of the Asp residues on the primary carbon atom of the epoxide ring of the target molecule, leading to a covalently bound ester intermediate (Arand, M. et al. (1996) J. Biol. Chem. 271:4223-4229; Rink, R. et al. (1997) J. Biol. Chem. 272:14650-14657; Argiriadi, M.A. et al. (2000) J. Biol. Chem. 275:15265-15270).

Enzymes involved in tyrosine catalysis

The degradation of the amino acid tyrosine, to either succinate and pyruvate or fumarate and acetoacetate, requires a large number of enzymes and generates a large number of intermediate compounds. In addition, many xenobiotic compounds may be metabolized using one or more reactions that are part of the tyrosine catabolic pathway. While the pathway has been studied primarily in bacteria, tyrosine degradation is known to occur in a variety of organisms and is likely to involve many of the same biological reactions.

The enzymes involved in the degradation of tyrosine to succinate and pyruvate (e.g., in *Arthrobacter* species) include 4-hydroxyphenylpyruvate oxidase, 4-hydroxyphenylacetate 3-hydroxylase, 3,4-dihydroxyphenylacetate 2,3-dioxygenase, 5-carboxymethyl-2-hydroxymuconic

semialdehyde dehydrogenase, *trans,cis*-5-carboxymethyl-2-hydroxymuconate isomerase, homoprotocatechuate isomerase/decarboxylase, *cis*-2-oxohept-3-ene-1,7-dioate hydratase, 2,4-dihydroxyhept-*trans*-2-ene-1,7-dioate aldolase, and succinic semialdehyde dehydrogenase.

The enzymes involved in the degradation of tyrosine to fumarate and acetoacetate (e.g., in *Pseudomonas* species) include 4-hydroxyphenylpyruvate dioxygenase, homogentisate 1,2-dioxygenase, maleylacetoacetate isomerase, and fumarylacetoacetase. 4-hydroxyphenylacetate 1-hydroxylase may also be involved if intermediates from the succinate/pyruvate pathway are accepted.

Additional enzymes associated with tyrosine metabolism in different organisms include 4-chlorophenylacetate-3,4-dioxygenase, aromatic aminotransferase, 5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase, 2-oxo-hept-3-ene-1,7-dioate hydratase, and 5-carboxymethyl-2-hydroxymuconate isomerase (Ellis, L.B.M. et al. (1999) Nucleic Acids Res. 27:373-376; Wackett, L.P. and Ellis, L.B.M. (1996) J. Microbiol. Meth. 25:91-93; and Schmidt, M. (1996) Amer. Soc. Microbiol. News 62:102).

In humans, acquired or inherited genetic defects in enzymes of the tyrosine degradation pathway may result in hereditary tyrosinemia. One form of this disease, hereditary tyrosinemia 1 (HT1) is caused by a deficiency in the enzyme fumarylacetoacetate hydrolase, the last enzyme in the pathway in organisms that metabolize tyrosine to fumarate and acetoacetate. HT1 is characterized by progressive liver damage beginning at infancy, and increased risk for liver cancer (Endo, F. et al. (1997) J. Biol. Chem. 272:24426-24432).

The discovery of new drug metabolizing enzymes, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of autoimmune/inflammatory, cell proliferative, neurological, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of drug metabolizing enzymes.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, drug metabolizing enzymes, referred to collectively as "DME" and individually as "DME-1," "DME-2," "DME-3," "DME-4," "DME-5," "DME-6," "DME-7," "DME-8," "DME-9," "DME-10," "DME-11," "DME-12," and "DME-13." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, c) a biologically

active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-13.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-13. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:14-26.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a

polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method

comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-13. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the

group consisting of SEQ ID NO:1-13. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional DME, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the

compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"DME" refers to the amino acid sequences of substantially purified DME obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of DME. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of DME either by directly interacting with DME or by acting on components of the biological pathway in which DME participates.

An "allelic variant" is an alternative form of the gene encoding DME. Allelic variants may

result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding DME include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as DME or a polypeptide with at least one functional characteristic of DME. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding DME, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding DME. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent DME. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of DME is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of DME. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of DME either by directly interacting with DME or by acting on components of the biological pathway in which DME participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind DME polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA;

RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic DME, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding DME or fragments of DME may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows

amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an

exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of DME or the polynucleotide encoding DME which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:14-26 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:14-26, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:14-26 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:14-26 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:14-26 and the region of SEQ ID NO:14-26 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-13 is encoded by a fragment of SEQ ID NO:14-26. A fragment of SEQ ID NO:1-13 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-13. For example, a fragment of SEQ ID NO:1-13 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-13. The precise length of a fragment of SEQ ID NO:1-13 and the region of SEQ ID NO:1-13 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer

to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example,

as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for

example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%.

Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of DME which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of DME which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of DME. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of DME.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an DME may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of DME.

"Probe" refers to nucleic acid sequences encoding DME, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge

MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing DME, nucleic acids encoding DME, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to

each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human drug metabolizing enzymes (DME), the polynucleotides encoding DME, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/inflammatory, cell proliferative, neurological, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homologs along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding

Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are drug metabolizing enzymes. SEQ ID NO:1 is 81% identical to rat cytochrome P450 (GenBank ID g203890) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.9e-225, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 is also 59% identical to rabbit cytochrome P450 2A10 (GenBank ID g165433) and 60% identical to rabbit cytochrome P450 2A11 (GenBank ID g165435), as determined by BLAST analysis, with probability scores of 1.6e-163 and 2.0e-163, respectively. SEQ ID NO:1 also contains a cytochrome P450 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a cytochrome P450.

SEQ ID NO:2 is 64% identical to human glutathione S-transferase omega (GenBank ID g8132762), as determined by BLAST analysis, with a probability score of 1.6e-83. SEQ ID NO:2 also contains a glutathione S-transferase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. Data from BLIMPS analysis provides further corroborative evidence that SEQ ID NO:2 is a glutathione S-transferase.

SEQ ID NO:3 is 76% identical to human heparan sulfate N-deacetylase/N-sulfotransferase 3 (GenBank ID g4322247), as determined by BLAST analysis, with a negligible probability score. SEQ ID NO:3 is also 78% identical to two other human heparan sulfate N-deacetylase/N-sulfotransferases (GenBank IDs g1036797 and g976372), based on BLAST analysis, also with negligible probability scores.

SEQ ID NO:4 is 58% identical to human N-acetylgalactosaminyltransferase (GenBank ID g2121220) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 8.7e-172, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 is also 57% and 46% identical to two other human N-acetylgalactosaminyltransferases (GenBank IDs g1934912 and g5834600,

respectively), based on BLAST analysis, with BLAST probability scores of 7.2e-168 and 1.4e-118, respectively. SEQ ID NO:4 also contains a glycosyltransferase active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS analysis provides further corroborative evidence that SEQ ID NO:4 is a N-acetylgalactosaminyltransferase.

SEQ ID NO:5 shares 98% overall identity to rat neuroligin 2 (GenBank ID g1145789) and 100% identity, over 550 amino acid residues, to a human neuroligin 3 isoform (GenBank ID g7960135), based on BLAST analysis. In both cases, the BLAST probability scores are negligible. The polynucleotide encoding the polypeptide of SEQ ID NO:5 (i.e., SEQ ID NO:18) and the polynucleotide encoding g1145789 (i.e., g1145788) share 78% overall identity and 92% identity within the respective coding regions, based on BLAST analysis. Neuroligins contain carboxylesterase-like domains although they are catalytically inactive. SEQ ID NO:5 contains a carboxylesterase-like domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS, PROFILESCAN, and BLIMPS analysis provides further corroborative evidence that SEQ ID NO:2 is a neuroligin.

SEQ ID NO:6 is 51% identical to human UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase (GenBank ID g971461) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.0e-146, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a glycosyl transferase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.)

SEQ ID NO:7 is 87% identical to murine carbonyl reductase (GenBank ID g1049108), a short-chain alcohol dehydrogenase, as determined by BLAST analysis. The BLAST probability score is 1.5e-128. SEQ ID NO:7 is also 85% identical to five human carbonyl reductases (GenBank IDs g179978, g181037, g3702679, g6693616, and g7768722), as determined by BLAST analysis, all with probability scores of 5.2e-126. SEQ ID NO:7 also contains a short-chain alcohol dehydrogenases domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:7 is a short-chain alcohol dehydrogenase.

SEQ ID NO:8 is 49% identical to Chinese hamster beta-1,6-N-acetylglucosaminyl transferase (GenBank ID g1698601), as determined by BLAST analysis, with a probability score of 1.1e-169. SEQ ID NO:8 is also 51% identical to a human beta-1,6-N-acetylglucosaminyl transferase (GenBank

ID g4545222), as determined by BLAST analysis, with a probability score of 2.0e-168.

SEQ ID NO:9 is 66% identical to an N-acetyltransferase from Pseudomonas aeruginosa (GenBank ID g9946340), as determined by BLAST analysis, with a probability score of 5.8e-56. SEQ ID NO:9 is also 35% identical to a murine (GenBank ID g309504) and a human (GenBank ID g1103904) N-acetyltransferase, as determined by BLAST analysis, with BLAST probability scores of 2.0e-21 and 6.9e-21, respectively. SEQ ID NO:9 also contains an acetyltransferase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. Data from BLIMPS analysis provide further corroborative evidence that SEQ ID NO:9 is an acetyltransferase.

SEQ ID NO:10 is 49% identical (over 525 amino acid residues) to a hamster carboxylesterase precursor (GenBank ID g2641986) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.9e-130, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 is also 48% identical to a rat carboxylesterase precursor (over 524 amino acid residues; GenBank ID g2641992) and 47% identical to human carboxylesterase hCE-2 (over 525 amino acid residues; GenBank ID g1407780), based on BLAST analysis, with BLAST probability scores of 3.1e-128 and 2.6e-124, respectively. SEQ ID NO:10 also contains a carboxylesterase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:10 is a carboxylesterase.

SEQ ID NO:11 is 62% identical (over 108 amino acid residues) to a Cu/Zn-superoxide dismutase from Caulobacter crescentus (GenBank ID g144283), as determined by BLAST analysis, with a probability score of 4.1e-32. SEQ ID NO:11 also contains a Cu/Zn-superoxide dismutase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:11 is a Cu/Zn-superoxide dismutase.

SEQ ID NO:12 is 86% identical, from residue M1 to residue F222, to four human glutathione S-transferases (GenBank IDs g306809, g306810, g259141, and g242749) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability scores are 6.3e-98, which indicate the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains a glutathione S-transferases domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis provides further

corroborative evidence that SEQ ID NO:12 is a glutathione S-transferase.

SEQ ID NO:13 is 55% identical, from residue R2 to residue V273, to Escherichia coli morphine-6-dehydrogenase (GenBank ID g1736409) as determined by BLAST analysis. The BLAST probability score is 8.0e-75. SEQ ID NO:2 also contains an aldo/keto reductase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:13 is an aldo/keto reductase. The algorithms and parameters for the analysis of SEQ ID NO:1-13 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:14-26 or that distinguish between SEQ ID NO:14-26 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the

polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX_gAAAAA_gBBBBB_1_N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, *gAAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, *gBBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, *gBBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses DME variants. A preferred DME variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the DME amino acid sequence, and which contains at least one functional or structural characteristic of DME.

The invention also encompasses polynucleotides which encode DME. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:14-26, which encodes DME. The polynucleotide sequences of SEQ ID NO:14-26, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding DME. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding DME. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:14-26 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:14-26. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of DME.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding DME. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding DME, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding DME over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding DME. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of DME.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding DME, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring DME, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode DME and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring DME under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding DME or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding DME and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode DME and DME derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding DME or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:14-26 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding DME may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode DME may be cloned in recombinant DNA molecules that direct expression of DME, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express DME.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter DME-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of DME, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding DME may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, DME itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis

may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of DME, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active DME, the nucleotide sequences encoding DME or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding DME. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding DME. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding DME and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding DME and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding DME. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV,

or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding DME. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding DME can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding DME into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of DME are needed, e.g. for the production of antibodies, vectors which direct high level expression of DME may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of DME. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of DME. Transcription of sequences encoding DME may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J.

6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding DME may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses DME in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of DME in cell lines is preferred. For example, sequences encoding DME can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g.,

Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding DME is inserted within a marker gene sequence, transformed cells containing sequences encoding DME can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding DME under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding DME and that express DME may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of DME using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on DME is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding DME include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding DME, or any fragments thereof, may be cloned into a vector

for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding DME may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode DME may be designed to contain signal sequences which direct secretion of DME through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding DME may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric DME protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of DME activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a

proteolytic cleavage site located between the DME encoding sequence and the heterologous protein sequence, so that DME may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled DME may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

DME of the present invention or fragments thereof may be used to screen for compounds that specifically bind to DME. At least one and up to a plurality of test compounds may be screened for specific binding to DME. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of DME, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which DME binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express DME, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing DME or cell membrane fractions which contain DME are then contacted with a test compound and binding, stimulation, or inhibition of activity of either DME or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with DME, either in solution or affixed to a solid support, and detecting the binding of DME to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

DME of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of DME. Such compounds may include agonists, antagonists, or partial or

inverse agonists. In one embodiment, an assay is performed under conditions permissive for DME activity, wherein DME is combined with at least one test compound, and the activity of DME in the presence of a test compound is compared with the activity of DME in the absence of the test compound. A change in the activity of DME in the presence of the test compound is indicative of a compound that modulates the activity of DME. Alternatively, a test compound is combined with an in vitro or cell-free system comprising DME under conditions suitable for DME activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of DME may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding DME or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding DME may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding DME can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding DME is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and

treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress DME, e.g., by secreting DME in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of DME and drug metabolizing enzymes. In addition, examples of tissues expressing DME can be found in Table 6. Therefore, DME appears to play a role in autoimmune/inflammatory, cell proliferative, neurological, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders. In the treatment of disorders associated with increased DME expression or activity, it is desirable to decrease the expression or activity of DME. In the treatment of disorders associated with decreased DME expression or activity, it is desirable to increase the expression or activity of DME.

Therefore, in one embodiment, DME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular

disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an endocrine disorder, such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism; disorders associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism; disorders associated with hyperthyroidism including thyrotoxicosis and its various

forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α-reductase, and gynecomastia; an eye disorder, such as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma, amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retrobulbar tumor, and chiasmal tumor; a metabolic disorder, such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, Menkes syndrome, occipital horn syndrome, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; hypocalcemia, hypophosphatemia, postpubescent cerebellar ataxia, and tyrosinemia, and a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, hereditary hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy,

hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas.

In another embodiment, a vector capable of expressing DME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified DME in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of DME may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME including, but not limited to, those listed above.

In a further embodiment, an antagonist of DME may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of DME. Examples of such disorders include, but are not limited to, those autoimmune/inflammatory, cell proliferative, neurological, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders described above. In one aspect, an antibody which specifically binds DME may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express DME.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding DME may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of DME including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of DME may be produced using methods which are generally known in the art. In particular, purified DME may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind DME. Antibodies to DME may also

be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with DME or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to DME have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of DME amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to DME may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce DME-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA

86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for DME may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between DME and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering DME epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for DME. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of DME-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple DME epitopes, represents the average affinity, or avidity, of the antibodies for DME. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular DME epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the DME-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of DME, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of DME-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g.,

Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding DME, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding DME. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding DME. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding DME may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides

brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in DME expression or regulation causes disease, the expression of DME from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in DME are treated by constructing mammalian expression vectors encoding DME and introducing these vectors by mechanical means into DME-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of DME include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). DME may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding DME from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to DME expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding DME under the control of an independent promoter or the retrovirus long

terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding DME to cells which have one or more genetic abnormalities with respect to the expression of DME. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding DME to target cells which have one or more genetic abnormalities with respect to the expression of DME. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing DME to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.

169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding DME to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for DME into the alphavirus genome in place of the capsid-coding region results in the production of a large number of DME-coding RNAs and the synthesis of high levels of DME in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of DME into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using

triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding DME.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding DME. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding DME. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not

limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased DME expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding DME may be therapeutically useful, and in the treatment of disorders associated with decreased DME expression or activity, a compound which specifically promotes expression of the polynucleotide encoding DME may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding DME is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding DME are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding DME. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable

for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of DME, antibodies to DME, and mimetics, agonists, antagonists, or inhibitors of DME.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising DME or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, DME or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et

al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example DME or fragments thereof, antibodies of DME, and agonists, antagonists or inhibitors of DME, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind DME may be used for the diagnosis of disorders characterized by expression of DME, or in assays to monitor patients being treated with DME or agonists, antagonists, or inhibitors of DME. Antibodies useful for diagnostic

purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for DME include methods which utilize the antibody and a label to detect DME in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring DME, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of DME expression. Normal or standard values for DME expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to DME under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of DME expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding DME may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of DME may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of DME, and to monitor regulation of DME levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding DME or closely related molecules may be used to identify nucleic acid sequences which encode DME. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding DME, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the DME encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:14-26 or from genomic sequences including promoters, enhancers, and introns of the DME gene.

Means for producing specific hybridization probes for DNAs encoding DME include the cloning of polynucleotide sequences encoding DME or DME derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes <u>in vitro</u> by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a

variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding DME may be used for the diagnosis of disorders associated with expression of DME. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic

nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an endocrine disorder, such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism; disorders associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism; disorders associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men,

Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α-reductase, and gynecomastia; an eye disorder, such as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma, amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retrobulbar tumor, and chiasmal tumor; a metabolic disorder, such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, Menkes syndrome, occipital horn syndrome, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; hypocalcemia, hypophosphatemia, postpubescent cerebellar ataxia, and tyrosinemia, and a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, hereditary hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas. The polynucleotide sequences encoding DME may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered DME expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding DME may be useful in assays that

detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding DME may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding DME in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of DME, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding DME, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding DME may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding DME, or a fragment of a polynucleotide complementary to the polynucleotide encoding DME, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or

quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding DME may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding DME are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of DME include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and

effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, DME, fragments of DME, or antibodies specific for DME may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms,

knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for DME to quantify the levels of DME expression. In one embodiment, the antibodies are used as elements on a microarray,

and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed.

(1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding DME may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding DME on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, DME, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a

solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between DME and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with DME, or fragments thereof, and washed. Bound DME is then detected by methods well known in the art. Purified DME can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding DME specifically compete with a test compound for binding DME. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with DME.

In additional embodiments, the nucleotide sequences which encode DME may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/254,308, U.S. Ser. No. 60/256,189, U.S. Ser. No. 60/257,713, U.S. Ser. No. 60/262,706, and U.S. Ser. No. 60/266,020, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA

purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows.. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently

analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:14-26. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative drug metabolizing enzymes were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode drug metabolizing enzymes, the encoded polypeptides were analyzed by querying against PFAM models for drug metabolizing enzymes. Potential drug metabolizing enzymes were also identified by homology to Incyte cDNA sequences that had been annotated as drug metabolizing enzymes. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to

correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in

Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of DME Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:14-26 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:14-26 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above. In this manner, SEQ ID NO:15 was mapped to chromosome 10 within the interval from 131.90 to 134.70 centiMorgans.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is

much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \times \text{minimum } \{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding DME are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding DME. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of DME Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham

Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:14-26 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated

using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the DME-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring DME. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of DME. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the DME-encoding transcript.

XII. Expression of DME

Expression and purification of DME is achieved using bacterial or virus-based expression systems. For expression of DME in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express DME upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of DME in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding DME by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, DME is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from DME at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman

Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified DME obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

DME function is assessed by expressing the sequences encoding DME at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of DME on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding DME and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding DME and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of DME Specific Antibodies

DME substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the DME amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-DME activity by, for example, binding the peptide or DME to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring DME Using Specific Antibodies

Naturally occurring or recombinant DME is substantially purified by immunoaffinity chromatography using antibodies specific for DME. An immunoaffinity column is constructed by covalently coupling anti-DME antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing DME are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of DME (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/DME binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and DME is collected.

XVI. Identification of Molecules Which Interact with DME

DME, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled DME, washed, and any wells with labeled DME complex are assayed. Data obtained using different concentrations of DME are used to calculate values for the number, affinity, and association of DME with the candidate molecules.

Alternatively, molecules interacting with DME are analyzed using the yeast two-hybrid

system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

DME may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of DME Activity

Cytochrome P450 activity of DME is measured using the 4-hydroxylation of aniline. Aniline is converted to 4-aminophenol by the enzyme, and has an absorption maximum at 630 nm (Gibson and Skett, supra). This assay is a convenient measure, but underestimates the total hydroxylation, which also occurs at the 2- and 3- positions. Assays are performed at 37°C and contain an aliquot of the enzyme and a suitable amount of aniline (approximately 2 mM) in reaction buffer. For this reaction, the buffer must contain NADPH or an NADPH-generating cofactor system. One formulation for this reaction buffer includes 85 mM Tris pH 7.4, 15 mM $MgCl_2$, 50 mM nicotinamide, 40 mg trisodium isocitrate, and 2 units isocitrate dehydrogenase, with 8 mg $NADP^+$ added to a 10 mL reaction buffer stock just prior to assay. Reactions are carried out in an optical cuvette, and the absorbance at 630 nm is measured. The rate of increase in absorbance is proportional to the enzyme activity in the assay. A standard curve can be constructed using known concentrations of 4-aminophenol.

1α,25-dihydroxyvitamin D 24-hydroxylase activity of DME is determined by monitoring the conversion of ^{3}H-labeled 1α,25-dihydroxyvitamin D (1α,25$(OH)_2$D) to 24,25-dihydroxyvitamin D (24,25$(OH)_2$D) in transgenic rats expressing DME. 1 μg of 1α,25$(OH)_2$D dissolved in ethanol (or ethanol alone as a control) is administered intravenously to approximately 6-week-old male transgenic rats expressing DME or otherwise identical control rats expressing either a defective variant of DME or not expressing DME. The rats are killed by decapitation after 8 hrs, and the kidneys are rapidly removed, rinsed, and homogenized in 9 volumes of ice-cold buffer (15 mM Tris-acetate (pH 7.4), 0.19 M sucrose, 2 mM magnesium acetate, and 5 mM sodium succinate). A portion (e.g., 3 ml) of each homogenate is then incubated with 0.25 nM 1α,25$(OH)_2$[1-^{3}H]D, with a specific activity of approximately 3.5 GBq/mmol, for 15 min at 37 °C under oxygen with constant shaking. Total lipids are extracted as described (Bligh, E.G. and W.J. Dyer (1959) Can. J. Biochem. Physiol. 37: 911-917) and the chloroform phase is analyzed by HPLC using a FINEPAK SIL column (JASCO, Tokyo, Japan) with an *n*-hexane/chloroform/methanol (10:2.5:1.5) solvent system at a flow rate of 1 ml/min. In the alternative, the chloroform phase is analyzed by reverse phase HPLC using a J SPHERE ODS-AM column (YMC Co. Ltd., Kyoto, Japan) with an acetonitrile buffer system (40 to 100%, in water, in 30 min) at a flow rate of 1 ml/min. The eluates are collected in fractions of 30

seconds (or less) and the amount of ^3H present in each fraction is measured using a scintillation counter. By comparing the chromatograms of control samples (i.e., samples comprising 1α,25-dihydroxyvitamin D or 24,25-dihydroxyvitamin D (24,25$(OH)_2$D), with the chromatograms of the reaction products, the relative mobilities of the substrate (1α,25$(OH)_2$[1-^{3}H]D) and product (24,25$(OH)_2$[1-^{3}H]D) are determined and correlated with the fractions collected. The amount of 24,25$(OH)_2$[1-^{3}H]D produced in control rats is subtracted from that of transgenic rats expressing DME. The difference in the production of 24,25$(OH)_2$[1-^{3}H]D in the transgenic and control animals is proportional to the amount of 25-hydrolase activity of DME present in the sample. Confirmation of the identity of the substrate and product(s) is confirmed by means of mass spectroscopy (Miyamoto, Y. et al. (1997) J. Biol. Chem. 272:14115-14119).

Flavin-containing monooxygenase activity of DME is measured by chromatographic analysis of metabolic products. For example, Ring, B.J. et al. (1999; Drug Metab. Dis. 27:1099-1103) incubated FMO in 0.1 M sodium phosphate buffer (pH 7.4 or 8.3) and 1 mM NADPH at 37°C, stopped the reaction with an organic solvent, and determined product formation by HPLC. Alternatively, activity is measured by monitoring oxygen uptake using a Clark-type electrode. For example, Ziegler, D.M. and Poulsen, L.L. (1978; Methods Enzymol. 52:142-151) incubated the enzyme at 37°C in an NADPH-generating cofactor system (similar to the one described above) containing the substrate methimazole. The rate of oxygen uptake is proportional to enzyme activity.

UDP glucuronyltransferase activity of DME is measured using a colorimetric determination of free amine groups (Gibson and Skett, supra). An amine-containing substrate, such as 2-aminophenol, is incubated at 37°C with an aliquot of the enzyme in a reaction buffer containing the necessary cofactors (40 mM Tris pH 8.0, 7.5 mM $MgCl_2$, 0.025% Triton X-100, 1 mM ascorbic acid, 0.75 mM UDP-glucuronic acid). After sufficient time, the reaction is stopped by addition of ice-cold 20% trichloroacetic acid in 0.1 M phosphate buffer pH 2.7, incubated on ice, and centrifuged to clarify the supernatant. Any unreacted 2-aminophenol is destroyed in this step. Sufficient freshly-prepared sodium nitrite is then added; this step allows formation of the diazonium salt of the glucuronidated product. Excess nitrite is removed by addition of sufficient ammonium sulfamate, and the diazonium salt is reacted with an aromatic amine (for example, N-naphthylethylene diamine) to produce a colored azo compound which can be assayed spectrophotometrically (at 540 nm, for example). A standard curve can be constructed using known concentrations of aniline, which will form a chromophore with similar properties to 2-aminophenol glucuronide.

Glutathione S-transferase activity of DME is measured using a model substrate, such as 2,4-dinitro-1-chlorobenzene, which reacts with glutathione to form a product, 2,4-dinitrophenyl-glutathione, that has an absorbance maximum at 340 nm. It is important to note that GSTs have differing substrate specificities, and the model substrate should be selected based on the substrate

preferences of the GST of interest. Assays are performed at ambient temperature and contain an aliquot of the enzyme in a suitable reaction buffer (for example, 1 mM glutathione, 1 mM dinitrochlorobenzene, 90 mM potassium phosphate buffer pH 6.5). Reactions are carried out in an optical cuvette, and the absorbance at 340 nm is measured. The rate of increase in absorbance is proportional to the enzyme activity in the assay.

N-acyltransferase activity of DME is measured using radiolabeled amino acid substrates and measuring radiolabel incorporation into conjugated products. Enzyme is incubated in a reaction buffer containing an unlabeled acyl-CoA compound and radiolabeled amino acid, and the radiolabeled acyl-conjugates are separated from the unreacted amino acid by extraction into n-butanol or other appropriate organic solvent. For example, Johnson, M. R et al. (1990; J. Biol. Chem. 266:10227-10233) measured bile acid-CoA:amino acid N-acyltransferase activity by incubating the enzyme with cholyl-CoA and ^{3}H-glycine or ^{3}H-taurine, separating the tritiated cholate conjugate by extraction into n-butanol, and measuring the radioactivity in the extracted product by scintillation. Alternatively, N-acyltransferase activity is measured using the spectrophotometric determination of reduced CoA (CoASH) described below.

N-acetyltransferase activity of DME is measured using the transfer of radiolabel from [^{14}C]acetyl-CoA to a substrate molecule (for example, see Deguchi, T. (1975) J. Neurochem. 24:1083-5). Alternatively, a spectrophotometric assay based on DTNB (5,5'-dithio-bis(2-nitrobenzoic acid; Ellman's reagent) reaction with CoASH may be used. Free thiol-containing CoASH is formed during N-acetyltransferase catalyzed transfer of an acetyl group to a substrate. CoASH is detected using the absorbance of DTNB conjugate at 412 nm (De Angelis, J. et al. (1997) J. Biol. Chem. 273:3045-3050). Enzyme activity is proportional to the rate of radioactivity incorporation into substrate, or the rate of absorbance increase in the spectrophotometric assay.

Protein arginine methyltransferase activity of DME is measured at 37 °C for various periods of time. S-adenosyl-L-[methyl-^{3}H]methionine ([^{3}H]AdoMet; specific activity = 75 Ci/mmol; NEN Life Science Products) is used as the methyl-donor substrate. Useful methyl-accepting substrates include glutathione S-transferase fibrillarin glycine-arginine domain fusion protein (GST-GAR), heterogeneous nuclear ribonucleoprotein (hnRNP), or hypomethylated proteins present in lysates from adenosine dialdehyde-treated cells. Methylation reactions are stopped by adding SDS-PAGE sample buffer. The products of the reactions are resolved by SDS-PAGE and visualized by fluorography. The presence of ^{3}H-labeled methyl-donor substrates is indicative of protein arginine methyltransferase activity of DME (Tang, J. et al. (2000) J. Biol. Chem. 275:7723-7730 and Tang, J. et al. (2000) J. Biol. Chem.275:19866-19876).

Catechol-*O*-methyltransferase activity of DME is measured in a reaction mixture consisting of 50 mM Tris-HCl (pH 7.4), 1.2 mM $MgCl_2$, 200 μM SAM (*S*-adenosyl-L-methionine) iodide

(containing 0.5 μCi of methyl-[H^3]SAM), 1 mM dithiothreitol, and varying concentrations of catechol substrate (e.g., L-dopa, dopamine, or DBA) in a final volume of 1.0 ml. The reaction is initiated by the addition of 250-500 μg of purified DME or crude DME-containing sample and performed at 37 °C for 30 min. The reaction is arrested by rapidly cooling on ice and immediately extracting with 7 ml of ice-cold n-heptane. Following centrifugation at 1000 x g for 10 min, 3-ml aliquots of the organic extracts are analyzed for radioactivity content by liquid scintillation counting. The level of catechol-associated radioactivity in the organic phase is proportional to the catechol-*O*-methyltransferase activity of DME (Zhu, B. T. and J. G. Liehr (1996) 271:1357-1363).

DHFR activity of DME is determined spectrophotometrically at 15 °C by following the disappearance of NADPH at 340 nm (ε_{340} = 11,800 $M^{-1} \cdot cm^{-1}$). The standard assay mixture contains 100 μM NADPH, 14 mM 2-mercaptoethanol, MTEN buffer (50 mM 2-morpholinoethanesulfonic acid, 25 mM tris(hydroxymethyl)aminomethane, 25 mM ethanolamine, and 100 mM NaCl, pH 7.0), and DME in a final volume of 2.0 ml. The reaction is started by the addition of 50 μM dihydrofolate (as substrate). The oxidation of NADPH to $NADP^+$ corresponds to the reduction of dihydrofolate in the reaction and is proportional to the amount of DHFR activity in the sample (Nakamura, T. and Iwakura, M. (1999) J. Biol. Chem. 274:19041-19047).

Aldo/keto reductase activity of DME is measured using the decrease in absorbance at 340 nm as NADPH is consumed. A standard reaction mixture is 135 mM sodium phosphate buffer (pH 6.2-7.2 depending on enzyme), 0.2 mM NADPH, 0.3 M lithium sulfate, 0.5-2.5 mg enzyme and an appropriate level of substrate. The reaction is incubated at 30°C and the reaction is monitored continuously with a spectrophotometer. Enzyme activity is calculated as mol NADPH consumed / mg of enzyme.

Alcohol dehydrogenase activity of DME is measured using the increase in absorbance at 340 nm as NAD^+ is reduced to NADH. A standard reaction mixture is 50 mM sodium phosphate, pH 7.5, and 0.25 mM EDTA. The reaction is incubated at 25°C and monitored using a spectrophotometer. Enzyme activity is calculated as mol NADH produced / mg of enzyme.

Carboxylesterase activity of DME is determined using 4-methylumbelliferyl acetate as a substrate. The enzymatic reaction is initiated by adding approximately 10 μl of DME-containing sample to 1 ml of reaction buffer (90 mM KH_2PO_4, 40 mM KCl, pH 7.3) with 0.5 mM 4-methylumbelliferyl acetate at 37°C. The production of 4-methylumbelliferone is monitored with a spectrophotometer (ε_{350} = 12.2 mM^{-1} cm^{-1}) for 1.5 min. Specific activity is expressed as micromoles of product formed per minute per milligram of protein and corresponds to the activity of DME in the sample (Evgenia, V. et al. (1997) J. Biol. Chem. 272:14769-14775).

In the alternative, the cocaine benzoyl ester hydrolase activity of DME is measured by incubating approximately 0.1 ml of DME and 3.3 mM cocaine in reaction buffer (50 mM NaH_2PO_4,

pH 7.4) with 1 mM benzamidine, 1 mM EDTA, and 1 mM dithiothreitol at 37°C. The reaction is incubated for 1 h in a total volume of 0.4 ml then terminated with an equal volume of 5% trichloroacetic acid. 0.1 ml of the internal standard 3,4-dimethylbenzoic acid (10 μg/ml) is added. Precipitated protein is separated by centrifugation at 12,000 × g for 10 min. The supernatant is transferred to a clean tube and extracted twice with 0.4 ml of methylene chloride. The two extracts are combined and dried under a stream of nitrogen. The residue is resuspended in 14% acetonitrile, 250 mM KH_2PO_4, pH 4.0, with 8 μl of diethylamine per 100 ml and injected onto a C18 reverse-phase HPLC column for separation. The column eluate is monitored at 235 nm. DME activity is quantified by comparing peak area ratios of the analyte to the internal standard. A standard curve is generated with benzoic acid standards prepared in a trichloroacetic acid-treated protein matrix (Evgenia, V. et al. (1997) J. Biol. Chem. 272:14769-14775).

In another alternative, DME carboxyl esterase activity against the water-soluble substrate para-nitrophenyl butyric acid is determined by spectrophotometric methods well known to those skilled in the art. In this procedure, the DME-containing samples are diluted with 0.5 M Tris-HCl (pH 7.4 or 8.0) or sodium acetate (pH 5.0) in the presence of 6 mM taurocholate. The assay is initiated by adding a freshly prepared para-nitrophenyl butyric acid solution (100 μg/ml in sodium acetate, pH 5.0). Carboxyl esterase activity is then monitored and compared with control autohydrolysis of the substrate using a spectrophotometer set at 405 nm (Wan, L. et al. (2000) J. Biol. Chem. 275:10041-10046).

Sulfotransferase activity of DME is measured using the incorporation of ^{35}S from [^{35}S]PAPS into a model substrate such as phenol (Folds, A. and J. L. Meek (1973) Biochim. Biophys. Acta 327:365-374). An aliquot of enzyme is incubated at 37°C with 1 mL of 10 mM phosphate buffer, pH 6.4, 50 mM phenol, and 0.4-4.0 mM [^{35}S] adenosine 3′-phosphate 5′-phosphosulfate (PAPS). After sufficient time for 5-20% of the radiolabel to be transferred to the substrate, 0.2 mL of 0.1 M barium acetate is added to precipitate protein and phosphate buffer. Then 0.2 mL of 0.1 M $Ba(OH)_2$ is added, followed by 0.2 mL $ZnSO_4$. The supernatant is cleared by centrifugation, which removes proteins as well as unreacted [^{35}S]PAPS. Radioactivity in the supernatant is measured by scintillation. The enzyme activity is determined from the number of moles of radioactivity in the reaction product.

Heparan sulfate 6-sulfotransferase activity of DME is measured in vitro by incubating a sample containing DME along with 2.5 μmol imidazole HCl (pH 6.8), 3.75 μg of protamine chloride, 25 nmol (as hexosamine) of completely desulfated and N-resulfated heparin, and 50 pmol (about 5 x 10^5 cpm) of [^{35}S]adenosine 3′-phosphate 5′-phosphosulfate (PAPS) in a final reaction volume of 50 μl at 37°C for 20 min. The reaction is stopped by immersing the reaction tubes in a boiling water bath for 1 min. 0.1 μmol (as glucuronic acid) of chondroitin sulfate A is added to the reaction mixture as a carrier. ^{35}S-labeled polysaccharides are precipitated with 3 volumes of cold ethanol

containing 1.3% potassium acetate and separated completely from unincorporated [^{35}S]PAPS and its degradation products by gel chromatography using desalting columns. One unit of enzyme activity is defined as the amount required to transfer 1 pmol of sulfate/min., determined by the amount of [^{35}S]PAPS incorporated into the precipitated polysaccharides (Habuchi, H. et al. (1995) J. Biol. Chem. 270:4172-4179).

In the alternative, heparan sulfate 6-sulfotransferase activity of DME is measured by extraction and renaturation of enzyme from gels following separation by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Following separation, the gel is washed with buffer (0.05 M Tris-HCl, pH 8.0), cut into 3-5 mm segments and subjected to agitation at 4 °C with 100 µl of the same buffer containing 0.15 M NaCl for 48 h. The eluted enzyme is collected by centrifugation and assayed for the sulfotransferase activity as described above (Habuchi, H. et al. (1995) J. Biol. Chem. 270:4172-4179).

In another alternative, DME sulfotransferase activity is determined by measuring the transfer of [^{35}S]sulfate from [^{35}S]PAPS to an immobilized peptide that represents the N-terminal 15 residues of the mature P-selectin glycoprotein ligand-1 polypeptide to which a C-terminal cysteine residue is added. The peptide spans three potential tyrosine sulfation sites. The peptide is linked via the cysteine residue to iodoacetamide-activated resin at a density of 1.5-3.0 µmol peptide/ml of resin. The enzyme assay is performed by combining 10 µl of peptide-derivitized beads with 2-20 µl of DME-containing sample in 40 mM Pipes (pH 6.8), 0.3 M NaCl, 20 mM $MnCl_2$, 50 mM NaF, 1% Triton X-100, and 1 mM 5'-AMP in a final volume of 130 µl. The assay is initiated by addition of 0.5 µCi of [^{35}S]PAPS (1.7 µM; 1 Ci = 37 GBq). After 30 min at 37°C, the reaction beads are washed with 6 M guanidine at 65°C and the radioactivity incorporated into the beads is determined by liquid scintillation counting. Transfer of [^{35}S]sulfate to the bead-associated peptide is measured to determine the DME activity in the sample. One unit of activity is defined as 1 pmol of product formed per min (Ouyang, Y.-B. et al. (1998) Biochemistry 95:2896-2901).

In another alternative, DME sulfotransferase assays are performed using [^{35}S]PAPS as the sulfate donor in a final volume of 30 µl, containing 50 mM Hepes-NaOH (pH 7.0), 250 mM sucrose, 1 mM dithiothreitol, 14 µM[^{35}S]PAPS (15 Ci/mmol), and dopamine (25 µM), *p*-nitrophenol (5 µM), or other candidate substrates. Assay reactions are started by the addition of a purified DME enzyme preparation or a sample containing DME activity, allowed to proceed for 15 min at 37°C, and terminated by heating at 100°C for 3 min. The precipitates formed are cleared by centrifugation. The supernatants are then subjected to the analysis of ^{35}S-sulfated product by either thin-layer chromatography or a two-dimensional thin layer separation procedure. Appropriate standards are run in parallel with the supernatants to allow the identification of the ^{35}S-sulfated products and determine the enzyme specificity of the DME-containing samples based on relative rates of migration of

reaction products (Sakakibara, Y. et al. (1998) J. Biol. Chem. 273:6242-6247).

Squalene epoxidase activity of DME is assayed in a mixture comprising purified DME (or a crude mixture comprising DME), 20 mM Tris-HCl (pH 7.5), 0.01 mM FAD, 0.2 unit of NADPH-cytochrome C (P-450) reductase, 0.01 mM [^{14}C]squalene (dispersed with the aid of 20 μl of Tween 80), and 0.2% Triton X-100. 1 mM NADPH is added to initiate the reaction followed by incubation at 37 °C for 30 min. The nonsaponifiable lipids are analyzed by silica gel TLC developed with ethyl acetate/benzene (0.5:99.5, v/v). The reaction products are compared to those from a reaction mixture without DME. The presence of 2,3(*S*)-oxidosqualene is confirmed using appropriate lipid standards (Sakakibara, J. et al. (1995) 270:17-20).

Epoxide hydrolase activity of DME is determined by following substrate depletion using gas chromatographic (GC) analysis of ethereal extracts or by following substrate depletion and diol production by GC analysis of reaction mixtures quenched in acetone. A sample containing DME or an epoxide hydrolase control sample is incubated in 10 mM Tris-HCl (pH 8.0), 1 mM ethylenediaminetetraacetate (EDTA), and 5 mM epoxide substrate (e.g., ethylene oxide, styrene oxide, propylene oxide, isoprene monoxide, epichlorohydrin, epibromohydrin, epifluorohydrin, glycidol, 1,2-epoxybutane, 1,2-epoxyhexane, or 1,2-epoxyoctane). A portion of the sample is withdrawn from the reaction mixture at various time points, and added to 1 ml of ice-cold acetone containing an internal standard for GC analysis (e.g., 1-nonanol). Protein and salts are removed by centrifugation (15 min, 4000 × g) and the extract is analyzed by GC using a 0.2 mm × 25-m CP-Wax57-CB column (CHROMPACK, Middelburg, The Netherlands) and a flame-ionization detector. The identification of GC products is performed using appropriate standards and controls well known to those skilled in the art. 1 unit of DME activity is defined as the amount of enzyme that catalyzes the production of 1 μmol of diol/min (Rink, R. et al. (1997) J. Biol. Chem. 272:14650-14657).

Aminotransferase activity of DME is assayed by incubating samples containing DME for 1 hour at 37°C in the presence of 1 mM L-kynurenine and 1 mM 2-oxoglutarate in a final volume of 200 μl of 150 mM Tris acetate buffer (pH 8.0) containing 70 μM PLP. The formation of kynurenic acid is quantified by HPLC with spectrophotometric detection at 330 nm using the appropriate standards and controls well known to those skilled in the art. In the alternative, L-3-hydroxykynurenine is used as substrate and the production of xanthurenic acid is determined by HPLC analysis of the products with UV detection at 340 nm. The production of kynurenic acid and xanthurenic acid, respectively, is indicative of aminotransferase activity (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

In another alternative, aminotransferase activity of DME is measured by determining the activity of purified DME or crude samples containing DME toward various amino and oxo acid

substrates under single turnover conditions by monitoring the changes in the UV/VIS absorption spectrum of the enzyme-bound cofactor, pyridoxal 5'-phosphate (PLP). The reactions are performed at 25°C in 50 mM 4-methylmorpholine (pH 7.5) containing 9 μM purified DME or DME containing samples and substrate to be tested (amino and oxo acid substrates). The half-reaction from amino acid to oxo acid is followed by measuring the decrease in absorbance at 360 nm and the increase in absorbance at 330 nm due to the conversion of enzyme-bound PLP to pyridoxamine 5' phosphate (PMP). The specificity and relative activity of DME is determined by the activity of the enzyme preparation against specific substrates (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937).

Superoxide dismutase activity of DME is assayed from cell pellets, culture supernatants, or purified protein preparations. Samples or lysates are resolved by electrophoresis on 15% non-denaturing polyacrylamide gels. The gels are incubated for 30 min in 2.5 mM nitro blue tetrazolium, followed by incubation for 20 min in 30 mM potassium phosphate, 30 mM TEMED, and 30 μM riboflavin (pH 7.8). Superoxide dismutase activity is visualized as white bands against a blue background, following illumination of the gels on a lightbox. Quantitation of superoxide dismutase activity is performed by densitometric scanning of the activity gels using the appropriate superoxide dismutase positive and negative controls (e.g., various amounts of commercially available E. coli superoxide dismutase (Harth, G. and Horwitz, M. A. (1999) J. Biol. Chem. 274:4281-4292).

XVIII. Identification of DME Inhibitors

Compounds to be tested are arrayed in the wells of a multi-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. DME activity is measured for each well and the ability of each compound to inhibit DME activity can be determined, as well as the dose-response profiles. This assay could also be used to identify molecules which enhance DME activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7483477	1	7483477CD1	14	7483477CB1
7485159	2	7485159CD1	15	7485159CB1
7485518	3	7485518CD1	16	7485518CB1
2860635	4	2860635CD1	17	2860635CB1
2530615	5	2530615CD1	18	2530615CB1
3883906	6	3883906CD1	19	3883906CB1
7473644	7	7473644CD1	20	7473644CB1
7485303	8	7485303CD1	21	7485303CB1
6970969	9	6970969CD1	22	6970969CB1
6538080	10	6538080CD1	23	6538080CB1
55048919	11	55048919CD1	24	55048919CB1
7485135	12	7485135CD1	25	7485135CB1
7684978	13	7684978CD1	26	7684978CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	7483477CD1	g203890	6.9e-225	Cytochrome P-450 [Rattus norvegicus]. Nef, P. et al. (1990) Olfactory-specific cytochrome P-450 (P-450olf1; IIG1): Gene structure and developmental regulation. J. Biol. Chem. 265:2903-2907.
2	7485159CD1	g8132762	1.6e-83	Glutathione transferase omega [Homo sapiens]. Board, P.G. et al. (2000) Identification, characterization, and crystal structure of the Omega class glutathione transferases. J. Biol. Chem. 275:24798-24806.
3	7485518CD1	g4322247	0.0	Heparan sulfate N-deacetylase/N-sulfotransferase 3. [Homo sapiens]. Aikawa, J.I. et al. (1999) Molecular cloning and expression of a third member of the heparan sulfate/heparin GlcNAc N-deacetylase/ N-sulfotransferase family. J. Biol. Chem. 274:2690-2695.
4	2860635CD1	g2121220	8.7e-172	Polypeptide GalNAc transferase-T4 [Mus musculus].
5	2530615CD1	g1145789	0.0	Neuroligin 2 [Rattus norvegicus]. Ichtchenko, K. et al. (1996) Structures, alternative splicing, and neurexin binding of multiple neuroligins. J. Biol. Chem. 271:2676-2682.
6	3883906CD1	g971461	3.0e-146	UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase [Homo sapiens]. White, T. et al. (1995) Purification and cDNA cloning of a human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. J. Biol. Chem. 270:24156-24165.
7	7473644CD1	g1049108	1.5e-128	Carbonyl reductase [Mus musculus]. Wei, J. (1996) Cloning a cDNA for carbonyl reductase (Cbr) from mouse cerebellum: murine genes that express cbr map to chromosomes 16 and 11. Genomics 34:147-148.

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
8	7485303CD1	g1698601	1.1e-169	Beta-1,6-N-acetylglucosaminyltransferase [Cricetulus griseus]. Weinstein, J. et al. (1996) A point mutation causes mistargeting of Golgi GlcNAc-TV in the Lec4A Chinese hamster ovary glycosylation mutant. J. Biol. Chem. 271:27462-27469.
9	6970969CD1	g9946340	5.8e-56	Probable N-acetyltransferase [Pseudomonas aeruginosa]. Stover, C.K. et al. (2000) Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen. Nature 406:959-964.
10	6538080CD1	g2641986	1.9e-130	Carboxylesterase precursor [Mesocricetus auratus].
11	55048919CD1	g144283	4.1e-32	Cu/Zn-superoxide dismutase [Caulobacter crescentus]. Steinman, H.M. and Ely, B. (1990) Copper-zinc superoxide dismutase of Caulobacter crescentus: Cloning, sequencing, and mapping of the gene and periplasmic location of the enzyme. J. Bacteriol. 172:2901-2910.
12	7485135CD1	g306810	6.3e-98	Glutathione S-transferase Ha subunit 1 (EC 2.5.1.18) [Homo sapiens]. Rhoads, D.M. et al. (1987) The basic glutathione S-transferases from human livers are products of separate genes. Biochem. Biophys. Res. Commun. 145:474-481.
13	7684978CD1	g1736409	8.0e-75	Morphine 6-dehydrogenase (EC 1.1.1.218) (Naloxone reductase) [Escherichia coli]. Itoh, T. et al. (1996) A 460-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 40.1-50.0 min region on the linkage map. DNA Res. 3:379-392.

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7483477CD1	494	S99, S106, S131, S144, Y155, S189, S208, S434, T309, T474, Y412	N206, N459	Signal Peptide: M1-G31	HMMER
					Cytochrome P450: P34-L491	HMMER_PFAM
					Transmembrane domain: G4-A23, N-terminus is extracellular	TMAP
					Cytochrome P450 cysteine heme-iron ligand signature cytochrome_p450.prf: P413-S461	PROFILESCAN
					CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC: PD000021: P34-F187, L244-G382, M140-K289, I335-Q420, K400-G479, E442-P467, P386-Y407	BLAST-PRODOM
					CYTOCHROME P450: DM00022|P24461|74-478: G74-G479	BLAST-DOMO
					E-class P450 group I signature: PR00463: R62-L81, A86-I107, T177-D195, N290-S307, L310-G336, D353-P371, L394-D418, F429-C439, C439-L450	BLIMPS-PRINTS
					P450 superfamily signature: PR00385: A301-M318, K319-N332, A354-I365, V430-C439, C439-F450	BLIMPS-PRINTS
					E-class P450 group II signature: PR00464: G121-G141, N290-M318, K319-G336, K348-M368, G388-K403, Y404-E419, N426-C439, C439-L462	BLIMPS-PRINTS
					Cytochrome P450 cysteine heme-iron ligand signature: F432-G441	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	7485159CD1	243	S28, S35, Y64, S152, S172,	N162	Glutathione S-transferases: I24-P211	HMMER-PFAM
					Glutathione S-transferase: PF00043: I72-K101	BLIMPS-PFAM
3	7485518CD1	873	S10, T42, T46, T67, T73, S74, S101, T240, S248, S292, T297, S328, S350, T356, T358, S371, S395, T447, S448, S499, S627, T630, T659, T720, S732, T750	N226, N342, N392, N658	Signal cleavage: M1-S29	SPSCAN
					Signal Peptide: M1-Y31	HMMER
					Transmembrane domains: K6-Y34, F277-I304, K498-N526, N-terminus is extracellular	TMAP
					SULFATE TRANSFERASE N-DEACETYLASE/N-SULFOTRANSFERASE SULFOTRANSFERASE GLUCOSAMINYL HEPARAN N-HSST DEACETYLASE/N-SULFOTRANSFERASE TRANSMEMBRANE: PD011877: V80-D593	BLAST-PRODOM
					SIGNAL-ANCHOR TRANSMEMBRANE DOMAIN: DM07899|P52848|1-881: R12-R873	BLAST-DOMO
4	2860635CD1	581	T5, T55, S97, S108, S158, T162, S178, T225, T289, S300, S318, T330, S343, S380, T411, Y433, S492, T524, S544, S555, T567		Signal cleavage: M1-G34	SPSCAN
					Signal Peptide Cleavage Site: A39	HMMER
					Glycosyl transferases: S139-F322	HMMER-PFAM
					Transmembrane domain: L14-R41, N-terminus is cytosolic	TMAP
					N-ACETYLGALACTOSAMINYLTRANSFERASE, TRANSFERASE POLYPEPTIDE, ACETYLGALACTOSAMINYLTRANSFERASE, UDP-GALNAC, POLYPEPTIDE GLYCOSYLTRANSFERASE: PD003162: W287-P443	BLAST-PRODOM
					ACETYLGALACTOSAMINYLTRANSFERASE; POLYPEPTIDE: DM03891|Q07537|32-558: P71-K577	BLAST-DOMO
					Cell attachment sequence (RGD): R220-D222	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	2530615CD1	835	S35, T156, S187, S223, S334, S404, T430, T450, T467, S524, T619, T640, Y672, S713, T832	N98, N136, N522, N823	Signal cleavage: M1-A14	SPSCAN
					Signal Peptide Cleavage Site: R16	HMMER
					Carboxylesterases: P42-W601	HMMER-PFAM
					Transmembrane domains: T192-R215, D671-Y699, N-terminus is cytosolic	TMAP
					Neuroligin signature: PR01090: R447-V461, P550-S571, Y580-V598, R670-Y699	BLIMPS-PRINTS
					Carboxylesterases type-B: BL00122: F69-A89, E139-P149, P178-Y188, V207-L222, G231-V271, L283-G292, W538-N548	BLIMPS-BLOCKS
					Lipolytic enzymes G-D-X (class of Carboxylesterases): BL01173: M180-T192, R258-V271	BLIMPS-BLOCKS
					Carboxylesterases type-B signatures (carboxylesterase_b.prf): N232-A287	PROFILESCAN
					ESTERASE HYDROLASE PRECURSOR SIGNAL GLYCOPROTEIN SERINE PROTEIN CARBOXYLESTERASE FAMILY, MULTIGENE: PD000169: L31-V338, C317-W601	BLAST-PRODOM
					CARBOXYLESTERASES TYPE-B DM00175|P23141|20-362: E151-E386, P42-K158	BLAST-DOMO
					Carboxylesterases type-B signature 2: E139-P149	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	3883906CD1	558	Y73, S75, Y86, S96, Y111, T142, S166, T208, S212, T279, S325, T367, T399, S411, S463, Y420, T449, S520, T540		Glycosyl transferases: S126-F304	HMMER-PFAM
					Transmembrane domains: M1-W25, V236-A253, P286-S302, N-terminus is non-cytoplasmic	TMAP
					N-ACETYLGALACTOSAMINYLTRANSFERASE; TRANSFERASE, POLYPEPTIDE, ACETYLGALACTOSAMINYLTRANSFERASE; UDP-GALNAC:POLYPEPTIDE GLYCOSYLTRANSFERASE: PD003162: K268-L431	BLAST-PRODOM
					ACETYLGALACTOSAMINYLTRANSFERASE; POLYPEPTIDE: DM03891|I37405|21-571: F68-Q553	BLAST-DOMO
7	7473644CD1	277	S30, T36, S56, S69, T101, T110, T162, T164, T179, S243	N137, N149	Short chain dehydrogenase: R6-D234	HMMER-PFAM
					Short-chain dehydrogenas: BL00061: G82-G92, K174-K211, K220-G229	BLIMPS-BLOCKS
					Short-chain dehydrogenases/reductases family signature (adh_short.prf): P131-C227	PROFILESCAN
					Alcohol dehydrogenase superfamily signature: PR00080: G82-I93, G133-L141, Y194-N213	BLIMPS-PRINTS
					Glucose/ribitol dehydrogenase (short-chain dehydrogenase) family signature: PR00081: G82-I93, P127-S143, Y194-N213, D219-A236, G237-L257, V7-D24	BLIMPS-PRINTS
					REDUCTASE CARBONYL NADPH NADPH-DEPENDENT OXIDOREDUCTASE NADP ACETYLATION INDUCIBLE 20-BETAHYDROXYSTEROID DEHYDROGENASE PD151158: H184-D234	BLAST-PRODOM
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY: DM00034|P16152|1-272: S2-V274	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Short-chain dehydrogenases/reductases family signature (Adh_Short): K181-A209	MOTIFS
8	7485303CD1	821	S19, S48, T57, S121, S200, Y216, S233, S277, T289, T293, T310, S333, S367, S412, T480, S528, T603, T644, T697, T732, T764	N158, N246, N267, N478, N629, N695, N704	Signal cleavage: M1-S17	SPSCAN
					Transmembrane domain: R335-R363	TMAP
					BETA1 ALPHA1 36-MANNOSYLGLYCOPROTEIN V ALPHAMANNOSIDE 6-N-ACETYL GLUCOSAMINYLTRANSFERASE 6-N-ACETYLGLUCOSAMINYLTRANSFERASE N-ACETYLGLUCOSAMINYLTRANSFERASE: PD025605: V180-C673, W712-L821, P788-V814	BLAST-PRODOM
9	6970969CD1	159	S36 S120		Acetyltransferase (GNAT) family: R58-Q138	HMMER-PFAM
					Transmembrane domain: G53-G77, N-terminus is non-cytoplasmic	TMAP
					Acetyltransferase (GNAT, GCN5-related): PF00583: I88-G98, A128-A137	BLIMPS-PFAM
					ACETYLTRANSFERASE; SPERMIDINE; SSAT; SPERMINE; DM08703|S61530|27-165: Y26-L153	BLAST-DOMO
10	6538080CD1	642	S49, T83, S136, S146, S222, T251, S350, S371, S395, S429, S462, S469, T486, T545, S552, S615	N348, N430, N580, N591	Signal cleavage: M1-L50	SPSCAN
					Carboxylesterases: G105-W606	HMMER-PFAM
					Transmembrane domains: T32-S56, T113-G135, K204-S220, D234-F250, V287-K314, L463-R491; N-terminus is non-cytoplasmic	TMAP
					Carboxylesterases type-B: BL00122: P206-F216, V235-F250, G259-V299, L311-G320, W559-N569, F124-P144, E186-P196	BLIMPS-BLOCKS
					Carboxylesterases type-B signatures (carboxylesterase_b.prf): N260-A315	PROFILESCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Cholinesterase signature: PR00878: W210-V239, L473-V481, H521-F533	BLIMPS-PRINTS
					ESTERASE HYDROLASE PRECURSOR SIGNAL GLYCOPROTEIN SERINE PROTEIN CARBOXYLESTERASE FAMILY MULTIGENE: PD000169: P142-L370, I87-W253, V452-W606, G248-Y500	BLAST-PRODOM
					CARBOXYLESTERASES TYPE-B: DM00175|P14943|2-326: P98-P418	BLAST-DOMO
					Carboxylesterases type-B serine active site: (Carboxylesterase_B_1): F280-G295	MOTIFS
					Carboxylesterases type-B signature 2 (Carboxylesterase_B_2): E186-P196	MOTIFS
11	55048919CD1	109	S69, S87		Copper/zinc superoxide dismutase (SODC): H2-R109	HMMER-PFAM
					Copper/Zinc superoxide domain: BL00087: L74-R109	BLIMPS-BLOCKS
					Copper/Zinc superoxide dismutase signatures (sod_cu_zn_2.prf): G79-I108	PROFILESCAN
					SUPEROXIDE DISMUTASE CUZN OXIDOREDUCTASE COPPER ZINC PRECURSOR SIGNAL PERIPLASMIC EXTRACELLULAR: PD000469: H2-I108 (p = 2.5e-09)	BLAST-PRODOM
					COPPER/ZINC SUPEROXIDE DISMUTASE DM00227|P20379|26-173: H2-R109	BLAST-DOMO
					Copper/Zinc superoxide dismutase signature 2 (Sod_Cu_Zn_2): G97-I108	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	7485135CD1	222	S37, S154, S202, S206, S212		Signal cleavage: M1-F30	SPSCAN
					Glutathione S-transferases: L7-P192	HMMER-PFAM
					Glutathione S-transferase signature: PF00043: Q54-R83	BLIMPS-PFAM
					GLUTATHIONE TRANSFERASE S-TRANSFERASE MULTIGENE FAMILY PROTEIN CLASS ALPHA S-CRYSTALLIN GST LYASE: PD000312: K87-Q199, K6-L102	BLAST-PRODOM
					GLUTATHIONE TRANSFERASE: DM00127|P08263|71-190: L72-P192	BLAST-DOMO
13	7684978CD1	273	T44, S84, S95, S139, T250, T265	N82	Signal cleavage: M1-G15	SPSCAN
					Aldo/keto reductase family: L45-L149, A185-R266	HMMER-PFAM
					Aldo/keto reductase family: BL00798: K5-R19, G67-Y79, A91-Y107, A205-D253, G38-G62	BLIMPS-BLOCKS
					Aldo/keto reductase family signatures (aldoketo_reductase_2.prf): V75-A164	PROFILESCAN
					Aldo-keto reductase signature: PR00069: R94-R112, F124-F141, L195-W219, G38-G62	BLIMPS-PRINTS
					ALDO/KETO REDUCTASE FAMILY: DM00192|P30863|1-256: S10-I257	BLAST-DOMO

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
14/ 7483477CB1/ 2004	1-647, 60-647, 77-650, 89-647, 101-636, 109-647, 117-647, 120-647, 123-647, 125-647, 130-647, 133-650, 141-593, 153-650, 157-647, 162-647, 163-647, 172-590, 172-600, 172-620, 172-640, 172-642, 172-643, 172-647, 173-643, 188-647, 189-647, 194-647, 200-647, 204-647, 212-647, 245-647, 256-647, 263-647, 267-647, 345-647, 358-647, 370-643, 389-647, 415-650, 428-647, 447-647, 458-647, 520-699, 520-862, 697-864, 698-864, 699-864, 700-1012, 710-864, 738-1011, 863-1012, 863-1173, 874-977, 874-1035, 1013-1173, 1013-1350, 1174-1350, 1174-1492, 1351-1492, 1351-1680, 1491-1681, 1493-1822, 1675-1702, 1675-1771, 1675-1775, 1675-1812, 1675-1815, 1675-1820, 1675-1823, 1681-1822, 1681-1823, 1681-2004, 1775-1821, 1823-2004
5/ 185159CB1/ 191	1-533, 135-357, 186-466, 324-595, 336-576, 336-592, 337-581, 338-1021, 340-610, 356-519, 356-528, 356-536, 356-549, 356-564, 356-569, 356-577, 356-600, 356-642, 356-813, 356-890, 356-937, 356-1021, 358-614, 366-939, 428-689, 429-1026, 450-689, 454-689, 465-689, 535-689, 546-1185, 585-936, 614-1191, 658-1092, 659-1168, 667-904, 707-1185, 730-956, 730-958, 731-1185, 740-1185, 745-1185, 767-1185, 768-1182, 790-898, 790-1166, 790-1182, 790-1185, 790-1188, 790-1189, 793-898, 804-1185, 806-1185, 825-1179, 837-1188
5/ 185518CB1/ 119	1-948, 792-1769, 891-1181, 951-1770, 1362-2015, 1781-2302, 1782-2328, 1946-2732, 1946-2750, 1951-2534, 2146-2752, 2529-3419
7/ 360635CB1/ 745	1-374, 198-649, 198-786, 198-821, 204-404, 204-458, 204-517, 204-525, 204-647, 204-674, 204-774, 204-795, 206-555, 206-706, 225-775, 229-890, 229-903, 241-629, 287-498, 329-598, 378-992, 396-917, 417-685, 418-948, 454-962, 489-1097, 491-826, 502-1070, 530-1070, 658-977, 791-1049, 791-1257, 791-1290, 791-1298, 791-1319, 791-1378, 791-1454, 802-1504, 910-1365, 937-1212, 970-1513, 975-1552, 1035-1288, 1050-1278, 1090-1366, 1090-1694, 1256-1490, 1280-1543, 1298-1558, 1298-1577, 1382-1627, 1426-1704, 1475-2148, 1479-1770, 1502-1786, 1564-1814, 1572-1861, 1589-2035, 1604-1977, 1616-2227, 1628-2284, 1631-2221, 1634-2234, 1648-1932, 1655-2039, 1661-1914, 1662-1935, 1662-1937, 1671-2039, 1674-2287,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1686-2319, 1688-2286, 1690-1966, 1695-1858, 1695-1885, 1695-2336, 1708-2389, 1713-2041, 1719-1927, 1720-2035, 1744-2389, 1745-2351, 1745-2394, 1750-2419, 1757-1994, 1758-2431, 1763-2345, 1776-2273, 1780-2420, 1784-2008, 1784-2032, 1784-2079, 1823-2467, 1826-2070, 1836-2103, 1860-2458, 1879-2471, 1880-2029, 1890-2168, 1893-2431, 1893-2523, 1908-2107, 1912-2499, 1912-2500, 1915-2528, 1924-2394, 1931-2579, 1934-2548, 1939-2183, 1945-2200, 1956-2630, 1977-2617, 1988-2606, 1989-2579, 1990-2213, 2002-2321, 2012-2685, 2036-2683, 2038-2670, 2038-2715, 2057-2669, 2061-2711, 2062-2652, 2067-2695, 2078-2713, 2092-2665, 2093-2314, 2101-2661, 2104-2516, 2107-2318, 2115-2403, 2116-2722, 2124-2434, 2125-2572, 2129-2279, 2129-2338, 2129-2705, 2131-2720, 2144-2723, 2151-2360, 2151-2424, 2152-2428, 2166-2648, 2172-2745, 2187-2638, 2196-2606, 2248-2492, 2249-2733, 2273-2733, 2275-2550, 2277-2740, 2282-2733, 2299-2736, 2308-2734, 2308-2745, 2312-2737, 2332-2734, 2333-2731, 2336-2727, 2337-2732, 2339-2731, 2340-2727, 2346-2733, 2358-2731, 2361-2734, 2368-2736, 2371-2734, 2385-2745, 2390-2731, 2398-2728, 2398-2730, 2417-2667, 2420-2731, 2470-2727, 2480-2734, 2536-2690, 2537-2745, 2543-2693, 2543-2718, 2543-2720, 2543-2731, 2543-2745, 2552-2687, 2552-2745
3/ 530615CB1/ 763	1-188, 1-550, 1-565, 213-569, 215-560, 228-569, 229-569, 290-567, 298-568, 302-369, 361-560, 365-569, 370-1103, 466-569, 496-703, 510-778, 511-702, 512-696, 538-703, 560-633, 560-635, 560-737, 560-872, 560-963, 560-985, 560-995, 560-1025, 560-1035, 560-1039, 560-1045, 560-1048, 560-1051, 560-1063, 560-1064, 560-1076, 560-1082, 560-1085, 560-1100, 560-1101, 560-1119, 560-1132, 560-1133, 560-1143, 560-1149, 560-1154, 560-1168, 560-1171, 560-1178, 560-1179, 560-1188, 560-1190, 560-1191, 560-1198, 560-1202, 560-1206, 560-1211, 560-1213, 560-1225, 560-1228, 560-1238, 560-1242, 560-1246, 560-1252, 560-1370, 561-693, 561-694, 561-697, 561-702, 561-703, 561-704, 561-879, 561-1013, 561-1051, 561-1266, 562-633, 563-1043, 563-1048, 564-809, 564-1034, 564-1203, 565-1103, 569-1228, 569-1287, 571-703, 572-1122, 579-945, 579-1026, 579-1094, 609-768, 631-1202, 634-1298, 675-1316, 679-1007, 679-1142, 680-1191, 684-993, 684-1193, 684-1275, 690-904, 690-911, 700-997, 721-1220, 732-1201, 747-1379, 757-1360, 761-1466, 785-1492, 861-1443, 905-1503, 976-1607, 998-1549, 998-1561, 1051-1584, 1124-1666, 1139-1721, 1166-1721, 1167-1728, 1190-1790, 1201-1790, 1223-1755, 1231-1748, 1234-1721, 1239-1588, 1257-1899, 1265-1790, 1266-1790,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1270-1790, 1272-1790, 1288-1721, 1301-1763, 1309-1716, 1311-1790, 1320-1790, 1329-1790, 1341-1626, 1355-1790, 1375-1686, 1389-2075, 1390-1721, 1394-1790, 1397-1790, 1406-1546, 1409-1790, 1414-2039, 1422-1733, 1431-2110, 1432-2068, 1490-2235, 1522-2114, 1538-2062, 1545-2010, 1557-2027, 1566-2051, 1574-2051, 1578-2033, 1579-2106, 1615-2185, 1645-2015, 1718-2185, 1782-2562, 1891-2575, 1901-2575, 1901-2576, 1905-2572, 1946-2508, 1964-2508, 2023-2508, 2033-2569, 2033-2575, 2033-2580, 2033-2581, 2121-2241, 2132-2487, 2145-2587, 2175-2581, 2218-2570, 2220-2508, 2389-2978, 2456-2978, 2471-3095, 2481-2979, 2506-3116, 2511-2979, 2538-2979, 2577-3088, 2603-3143, 2634-2943, 2648-3086, 2648-3105, 2648-3141, 2648-3148, 2648-3208, 2648-3217, 2648-3222, 2648-3227, 2702-2958, 2703-3218, 2783-3043, 2803-3071, 2816-3083, 2831-3060, 2862-3088, 2880-3139, 2898-3476, 2942-3214, 2942-3471, 2992-3254, 3011-3273, 3054-3488, 3118-3631, 3128-3384, 3139-3403, 3145-3603, 3254-3543, 3287-3524, 3347-3584, 3347-3587, 3359-3598, 3359-3701, 3369-3629, 3373-3957, 3394-3634, 3425-3977, 3444-3686, 3444-3744, 3447-3666, 3462-3723, 3467-3722, 3539-3798, 3543-3794, 3560-3772, 3564-3818, 3565-3832, 3595-3827, 3617-3897, 3629-3930, 3668-3752, 3686-4009, 3701-3891, 3711-3981, 3714-3946, 3748-3961, 3774-4037, 3778-4019, 3778-4056, 3783-3994, 3788-4030, 3788-4042, 3794-4085, 3834-4084, 3877-4150, 3948-4208, 3974-4272, 3985-4217, 4002-4297, 4009-4197, 4034-4305, 4045-4243, 4097-4736, 4116-4405, 4142-4622, 4151-4401, 4162-4431, 4162-4634, 4171-4582, 4175-4763, 4179-4592, 4179-4593, 4182-4436, 4182-4727, 4182-4763, 4183-4481, 4184-4390, 4184-4489, 4187-4681, 4190-4608, 4190-4639, 4190-4641, 4202-4457, 4203-4452, 4204-4481, 4204-4482, 4205-4392, 4212-4465, 4217-4530, 4220-4454, 4220-4759, 4230-4466, 4232-4409, 4252-4515, 4262-4763, 4269-4545, 4269-4557, 4274-4749, 4277-4763, 4284-4702, 4293-4763, 4300-4738, 4306-4763, 4307-4763, 4316-4591, 4322-4754, 4323-4752, 4325-4556, 4355-4763, 4366-4745, 4367-4612, 4367-4701, 4367-4745, 4376-4621, 4376-4735, 4383-4699, 4400-4748, 4407-4747, 4409-4745, 4412-4744, 4418-4745, 4443-4621, 4494-4730, 4587-4758, 4610-4763, 4613-4745

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
19/3883906CB1/ 2970	1-249, 1-532, 10-129, 10-339, 21-602, 27-475, 75-332, 331-593, 331-861, 602-652, 610-859, 610-1102, 610-1105, 944-1568, 945-1252, 949-1204, 949-1511, 992-1608, 992-1609, 992-1628, 992-1669, 992-1683, 992-1715, 994-1291, 1098-1308, 1098-1736, 1098-1738, 1108-1566, 1186-1732, 1200-1472, 1217-1455, 1254-1545, 1267-1839, 1302-1941, 1465-1679, 1596-1955, 1618-1785, 1623-2316, 1650-2182, 1680-2160, 1680-2186, 1785-2144, 1814-2276, 1814-2289, 1834-2289, 1836-2082, 1856-2054, 1856-2313, 1856-2400, 1911-2150, 1912-2574, 1961-2293, 1965-2254, 1970-2254, 1973-2403, 2116-2434, 2193-2529, 2333-2902, 2343-2901, 2346-2902, 2352-2893, 2371-2916, 2483-2920, 2483-2945, 2483-2947, 2491-2920, 2499-2946, 2505-2941, 2515-2938, 2565-2914, 2577-2920, 2577-2925, 2604-2912, 2607-2937, 2649-2929, 2688-2854, 2688-2905, 2758-2970, 2786-2920, 2868-2970
)/7473644CB1/ 34	1-834
L/ 185303CB1/ 166	1-194, 87-645, 310-422, 423-1118, 784-1118, 846-1379, 1119-1384, 1145-1379, 1213-1384, 1251-1515, 1385-2466, 1514-1767, 1514-1821, 1516-1575, 1516-1720, 1516-1906, 1525-1821, 1752-2113, 1793-2037, 1817-1937
2/ 270969CB1/ 144	1-682, 21-561, 21-577, 21-609, 21-615, 21-647, 21-686, 21-692, 21-698, 21-730, 21-749, 21-768, 21-882, 22-703, 23-746, 49-759, 67-862, 98-747, 126-852, 146-988, 170-572, 178-893, 283-1195, 360-967, 367-926, 420-1117, 449-1233, 462-1188, 465-1195, 466-1164, 506-1233, 606-1135, 696-983, 911-1441, 990-1444, 992-1444, 1004-1444
3/6538080CB1/ 232	1-900, 451-1056, 622-870, 626-870, 640-718, 640-871, 641-870, 643-871, 701-1379, 701-1474, 746-871, 805-871, 845-871, 1137-1474, 1137-1700, 1137-1748, 1141-1474, 1463-2232, 1508-2232, 1546-2232, 1625-1922, 1625-2232, 1631-2232, 1634-2232, 1649-2232
24/55048919CB1/ 816	1-540, 76-816
25/7485135CB1/ 922	1-669, 274-416, 415-547, 528-922
26/7684978CB1/ 1084	1-494, 1-497, 1-569, 1-759, 4-115, 427-1041, 427-1073, 427-1077, 427-1080, 427-1082, 427-1084

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
14	7483477CB1	BRAITDR02
15	7485159CB1	BRSTNOT16
16	7485518CB1	SINTNON02
17	2860635CB1	CERVNOT01
18	2530615CB1	KIDEUNE02
19	3883906CB1	PROSTUS25
21	7485303CB1	BRAIFER06
22	6970969CB1	BMARUNR02
23	6538080CB1	OVARDIR01
24	55048919CB1	ADMEDNV37
25	7485135CB1	ADRENOT07
26	7684978CB1	BRABDIK02

Table 6

Library	Vector	Library Description
ADMEDNV37	pCR2-TopoTA	Library was constructed using pooled cDNA from 111 different donors. cDNA was generated using mRNA isolated from pooled skeletal muscle tissue removed from 10 Caucasian male and female donors, ages 21-57, who died from sudden death; from pooled thymus tissue removed from 9 Caucasian male and female donors, ages 18-32, who died from sudden death; from pooled fetal liver tissue removed from 32 Caucasian male and female fetuses, ages 18-24 weeks, who died from spontaneous abortions; from pooled fetal kidney tissue removed from 59 Caucasian male and female fetuses, ages 20-33 weeks, who died from spontaneous abortions; and from fetal brain tissue removed from a 23-week-old Caucasian male fetus who died from fetal demise.
ADRENOT07	pINCY	Library was constructed using RNA isolated from adrenal tissue removed from a 61-year-old female during a bilateral adrenalectomy. Patient history included an unspecified disorder of the adrenal glands.
BMARUNR02	PIGEN	This random primed library was constructed using RNA isolated from an untreated SH-SY5Y cell line derived from bone marrow neuroblastoma tumor cells removed from a 4-year-old Caucasian female.
BRABDIK02	PSPORT1	This amplified and normalized library was constructed using pooled cDNA from three different donors. cDNA was generated using mRNA isolated from diseased vermis tissue removed from a 79-year-old Caucasian female (donor A) who died from pneumonia, an 83-year-old Caucasian male (donor B) who died from congestive heart failure, and an 87-year-old Caucasian female (donor C) who died from esophageal cancer. Pathology indicated severe Alzheimer's disease in donors A & B and moderate Alzheimer's disease in donor C. Patient history included glaucoma, pseudophakia, gastritis with gastrointestinal bleeding, peripheral vascular disease, chronic obstructive pulmonary disease, seizures, tobacco abuse in remission, and transitory ischemic attacks in donor A; Parkinson's disease and atherosclerosis in donor B; hypertension, coronary artery disease, cerebral vascular accident, and hypothyroidism in donor C. Family history included Alzheimer's disease in the mother and sibling(s) of donor A. Independent clones from this amplified library were normalized in one round using conditions adapted Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.

Table 6 (cont.)

Library	Vector	Library Description
BRAIFER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRAITDR02	PCDNA2.1	This random primed library was constructed using RNA isolated from allocortex, neocortex, anterior and frontal cingulate tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRSTNOT16	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 59-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive lobular carcinoma with extension into ducts. Patient history included liver cirrhosis, esophageal ulcer, hyperlipidemia, and neuropathy.
CERVNOT01	PSPORT1	Library was constructed using RNA isolated from the uterine cervical tissue of a 35-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. Pathology indicated mild chronic cervicitis. Family history included atherosclerotic coronary artery disease and type II diabetes.
KIDEUNE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from an untreated transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue (Invitrogen). The cells were transformed with adenovirus 5 DNA.

Table 6 (cont.)

Library	Vector	Library Description
OVARDIR01	PCDNA2.1	This random primed library was constructed using RNA isolated from right ovary tissue removed from a 45-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, vaginal suspension and fixation, and incidental appendectomy. Pathology indicated stromal hyperthecosis of the right and left ovaries. Pathology for the matched tumor tissue indicated a dermoid cyst (benign cystic teratoma) in the left ovary. Multiple (3) intramural leiomyomata were identified. The cervix showed squamous metaplasia. Patient history included metrorrhagia, female stress incontinence, alopecia, depressive disorder, pneumonia, normal delivery, and deficiency anemia. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, and primary tuberculous complex.
PROSTUS25	pINCY	This subtracted prostate tumor tissue library was constructed using 2.36 million clones from a prostate tumor tissue library and was subjected to two rounds of subtraction hybridization with 2.36 million clones from an untreated prostate epithelial cell tissue library. The starting library for subtraction was constructed using RNA isolated from prostate tumor tissue removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3) involving the left and right sides of the prostate peripherally with invasion of the capsule. Adenofibromatous hyperplasia was present. Patient history included elevated prostate-specific antigen (PSA), diverticulitis of colon, asbestosis, and thrombophlebitis. Family history included benign hypertension multiple myeloma, hyperlipidemia, and rheumatoid arthritis. The hybridization probe for subtraction was derived from a similarly constructed library using RNA isolated from untreated prostate epithelial cell tissue from a different donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al., Genome Research (1996) 6:791.

Table 6 (cont.)

Library	Vector	Library Description
SINTNON02	pINCY	This normalized small intestine tissue library was constructed from 1.84 million independent clones from a pooled small intestine tissue library. Starting RNA was made from pooled cDNA from six different donors. cDNA was generated using mRNA isolated from small intestine tissue removed from a Caucasian male fetus (donor A) who died from fetal demise; from small intestine tissue removed from a 8-year-old Black male (donor B) who died from anoxia; from small intestine tissue removed from a 13-year-old Caucasian male (donor C) who died from a gunshot wound to the head; from jejunum and duodenum tissue removed from the small intestine of a 16-year-old Caucasian male (donor D) who died from head trauma; from ileum tissue removed from an 8-year-old Caucasian female (donor E) who died from head trauma; and from small intestine tissue removed from a 15-year-old Caucasian female who died from a closed head injury. Serologies were negative for donors A, B, C and D. Donors E and F had serologies positive for cytomegalovirus (CMV). Donor B's medications included DDAVP, Versed, and labetalol. Donor C's previous surgeries included a hernia repair. The patient was not taking any medications. Family history included diabetes in the grandparent(s). Donor D's history included a kidney infection three years prior to death, marijuana use, and tobacco use. Donor E's history included migraine headaches and urinary tract infection. Previous surgeries included an adenotonsillectomy. Patient medications included Dilantin (phenytoin), Ancef (cephalosporin), and Zantac (ranitidine). Donor F's history included seasonal allergies and marijuana use. Patient medications included Dopamine and Neo-Synephrine. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-4 and SEQ ID NO:6-13,

c) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to the amino acid sequence of SEQ ID NO:5,

d) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, and

e) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-13.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant

polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-13.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-26,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:14-19, SEQ ID NO:20-26,

c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 95% identical to the polynucleotide sequence of SEQ ID NO:18,

d) a polynucleotide complementary to a polynucleotide of a),

e) a polynucleotide complementary to a polynucleotide of b),

f) a polynucleotide complementary to a polynucleotide of c), and

g) an RNA equivalent of a)-f).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-13.

19. A method for treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional DME, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of DME in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of DME in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of DME in

a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-13, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-13.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

<110> INCYTE GENOMICS, INC.
SANJANWALA, Madhu M.
YAO, Monique G.
AU-YOUNG, Janice
BAUGHN, Mariah R.
ARVIZU, Chandra
RING, Huijun Z.
LEE, Ernestine A.
DING, Li
HAFALIA, April J.A.
TANG, Y. Tom
HUE, Henry
TRIBOULEY, Catherine M.
LU, Dyung Aina M.
LAL, Preeti G.
WARREN, Bridget A.
YANG, Junming
WALIA, Narinder K.
NGUYEN, Danniel B.
GANDHI, Ameena R.
LU, Yan
ISON, Craig H.

<120> DRUG METABOLIZING ENZYMES

<130> PI-0313 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/254,308; 60/256,189; 60/257,713; 60/262,706
60/266,020
<151> 2000-12-08; 2000-12-15; 2000-12-21; 2001-01-19
2001-02-02

<160> 26
<170> PERL Program

<210> 1
<211> 494
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7483477CD1

<400> 1
Met Glu Leu Gly Gly Ala Val Thr Ile Phe Leu Ala Leu Arg Leu
1 5 10 15
Ser Cys Leu Leu Ile Leu Ile Ala Trp Lys Arg Met Asp Lys Ala
20 25 30
Gly Lys Leu Pro Pro Gly Pro Thr Pro Ile Pro Phe Leu Gly Asn
35 40 45
Leu Leu Gln Val Arg Thr Asp Ala Thr Phe Gln Ser Phe Met Lys

```
                50                  55                  60
Leu Arg Glu Lys Tyr Ser Pro Val Phe Thr Val Tyr Met Gly Pro
                65                  70                  75
Arg Pro Val Val Val Leu Cys Gly His Glu Ala Val Lys Glu Ala
                80                  85                  90
Leu Ile Asp Gln Ala Asp Glu Phe Ser Gly Arg Gly Glu Leu Ala
                95                 100                 105
Ser Ile Lys Gln Asn Phe Gln Gly His Gly Val Ala Leu Ala Asn
               110                 115                 120
Gly Glu Arg Trp Arg Ile Leu Arg Arg Phe Ser Leu Thr Ile Leu
               125                 130                 135
Arg Asp Phe Gly Met Gly Lys Gln Ser Ile Lys Glu Arg Ile Gln
               140                 145                 150
Glu Glu Ala Ser Tyr Leu Leu Glu Glu Phe Gln Lys Thr Lys Gly
               155                 160                 165
Ala Pro Ile Asp Pro Ile Phe Leu Leu Ser Arg Thr Val Ser Asn
               170                 175                 180
Val Ile Ser Ser Val Val Phe Arg Ser Arg Phe Asp Tyr Glu Asp
               185                 190                 195
Lys Gln Phe Leu Asn Leu Leu Arg Leu Ile Asn Glu Ser Phe Ile
               200                 205                 210
Glu Met Ser Thr Pro Trp Ala Gln Leu Tyr Asp Met Tyr Ser Gly
               215                 220                 225
Ile Met Gln Tyr Leu Pro Gly Arg His Asn Leu Ile Tyr Tyr Leu
               230                 235                 240
Val Glu Glu Leu Lys Asp Phe Ile Ala Ser Arg Val Lys Ile Asn
               245                 250                 255
Glu Ala Ser Phe Asp Pro Gln Asn Pro Arg Asp Phe Ile Asp Cys
               260                 265                 270
Phe Leu Ile Lys Met His Gln Asp Lys Asn Asn Pro Arg Thr Glu
               275                 280                 285
Phe Asn Leu Lys Asn Leu Val Leu Thr Thr Leu Asn Leu Phe Phe
               290                 295                 300
Ala Gly Thr Glu Thr Val Ser Ser Thr Leu Arg Tyr Gly Phe Leu
               305                 310                 315
Leu Leu Met Lys His Pro Glu Val Glu Ala Lys Ile His Glu Glu
               320                 325                 330
Ile Asn Gln Val Ile Gly Pro His Arg Leu Pro Arg Val Asp Asp
               335                 340                 345
Arg Val Lys Met Pro Tyr Thr Asp Ala Val Ile His Glu Ile Gln
               350                 355                 360
Arg Leu Val Asp Ile Val Pro Met Gly Val Pro His Asn Leu Ile
               365                 370                 375
Arg Asp Thr Gln Phe Arg Gly Tyr Leu Leu Pro Lys Gly Thr Asp
               380                 385                 390
Val Phe Pro Leu Leu Gly Ser Val Leu Lys Asp Pro Lys Tyr Phe
               395                 400                 405
Arg Tyr Pro Asp Ala Phe Tyr Pro Gln His Phe Leu Asp Glu Gln
               410                 415                 420
Gly Arg Phe Lys Lys Asn Glu Ala Phe Val Pro Phe Ser Ser Gly
               425                 430                 435
Lys Arg Ile Cys Leu Gly Glu Ala Met Asp Arg Met Glu Leu Phe
               440                 445                 450
Leu Tyr Phe Thr Ser Thr Leu Gln Asn Phe Ser Leu His Ser Leu
               455                 460                 465
Val Pro Pro Val Asp Ile Asp Ile Thr Pro Lys Leu Ser Gly Phe
```

```
                470                 475                 480
Gly Asn Ile Pro Pro Thr Tyr Glu Leu Cys Leu Val Ala Arg
                485                 490

<210> 2
<211> 243
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485159CD1

<400> 2
Met Ser Gly Asp Ala Thr Arg Thr Leu Gly Lys Gly Ser Gln Pro
  1               5                  10                  15
Pro Gly Pro Val Pro Glu Gly Leu Ile Arg Ile Tyr Ser Met Arg
                 20                  25                  30
Phe Cys Pro Tyr Ser His Arg Thr Arg Leu Val Leu Lys Ala Lys
                 35                  40                  45
Asp Ile Arg His Glu Val Val Asn Ile Asn Leu Arg Asn Lys Pro
                 50                  55                  60
Glu Trp Tyr Tyr Thr Lys His Pro Phe Gly His Ile Pro Val Leu
                 65                  70                  75
Glu Thr Ser Gln Cys Gln Leu Ile Tyr Glu Ser Val Ile Ala Cys
                 80                  85                  90
Glu Tyr Leu Asp Asp Ala Tyr Pro Gly Arg Lys Leu Phe Pro Tyr
                 95                 100                 105
Asp Pro Tyr Glu Arg Ala Arg Gln Lys Met Leu Leu Glu Leu Phe
                110                 115                 120
Cys Lys Val Pro His Leu Thr Lys Glu Cys Leu Val Ala Leu Arg
                125                 130                 135
Cys Gly Arg Glu Cys Thr Asn Leu Lys Ala Ala Leu Arg Gln Glu
                140                 145                 150
Phe Ser Asn Leu Glu Glu Ile Leu Glu Tyr Gln Asn Thr Thr Phe
                155                 160                 165
Phe Gly Gly Thr Cys Ile Ser Met Ile Asp Tyr Leu Leu Trp Pro
                170                 175                 180
Trp Phe Glu Arg Leu Asp Val Tyr Gly Ile Leu Asp Cys Val Ser
                185                 190                 195
His Thr Pro Ala Leu Arg Leu Trp Ile Ser Ala Met Lys Trp Asp
                200                 205                 210
Pro Thr Val Cys Ala Leu Leu Met Asp Lys Ser Ile Phe Gln Gly
                215                 220                 225
Phe Leu Asn Leu Tyr Phe Gln Asn Asn Pro Asn Ala Phe Asp Phe
                230                 235                 240
Gly Leu Cys

<210> 3
<211> 873
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
```

<223> Incyte ID No: 7485518CD1

<400> 3

```
Met Asn Leu Ile Val Lys Leu Arg Arg Ser Phe Arg Thr Leu Ile
  1               5                  10                  15
Val Leu Leu Ala Thr Phe Cys Leu Val Ser Ile Val Ile Ser Ala
             20                  25                  30
Tyr Phe Leu Tyr Ser Gly Tyr Lys Gln Glu Met Thr Leu Ile Glu
             35                  40                  45
Thr Thr Ala Glu Ala Glu Cys Thr Asp Ile Lys Ile Leu Pro Tyr
             50                  55                  60
Arg Ser Met Glu Leu Lys Thr Val Lys Pro Ile Asp Thr Ser Lys
             65                  70                  75
Thr Asp Pro Thr Val Leu Leu Phe Val Glu Ser Gln Tyr Ser Gln
             80                  85                  90
Leu Gly Gln Asp Ile Ile Ala Ile Leu Glu Ser Ser Arg Phe Gln
             95                 100                 105
Tyr His Met Val Ile Ala Pro Gly Lys Gly Asp Ile Pro Pro Leu
            110                 115                 120
Thr Asp Asn Gly Lys Gly Lys Tyr Thr Leu Val Ile Tyr Glu Asn
            125                 130                 135
Ile Leu Lys Tyr Val Ser Met Asp Ser Trp Asn Arg Glu Leu Leu
            140                 145                 150
Glu Lys Tyr Cys Val Glu Tyr Ser Val Ser Ile Ile Gly Phe His
            155                 160                 165
Lys Ala Asn Glu Asn Ser Leu Pro Ser Thr Gln Leu Lys Gly Phe
            170                 175                 180
Pro Leu Asn Leu Phe Asn Asn Leu Ala Leu Lys Asp Cys Phe Val
            185                 190                 195
Asn Pro Gln Ser Pro Leu Leu His Ile Thr Lys Ala Pro Lys Val
            200                 205                 210
Glu Lys Gly Pro Leu Pro Gly Glu Asp Trp Thr Ile Phe Gln Tyr
            215                 220                 225
Asn His Ser Thr Tyr Gln Pro Val Leu Leu Thr Glu Leu Gln Thr
            230                 235                 240
Glu Lys Ser Leu Ser Ser Leu Ser Ser Lys Thr Leu Phe Ala Thr
            245                 250                 255
Val Ile Gln Asp Leu Gly Leu His Asp Gly Ile Gln Arg Val Leu
            260                 265                 270
Phe Gly Asn Asn Leu Asn Phe Trp Leu His Lys Leu Ile Phe Ile
            275                 280                 285
Asp Ala Ile Ser Phe Leu Ser Gly Lys Arg Leu Thr Leu Ser Leu
            290                 295                 300
Asp Arg Tyr Ile Leu Val Asp Ile Asp Asp Ile Phe Val Gly Lys
            305                 310                 315
Glu Gly Thr Arg Met Asn Val Lys Asp Val Lys Val Ser Leu Phe
            320                 325                 330
Glu Thr Gln Asn Leu Leu Arg Thr Gln Val Ala Asn Phe Thr Phe
            335                 340                 345
Asn Leu Gly Phe Ser Gly Lys Phe Tyr His Thr Gly Thr Glu Glu
            350                 355                 360
Glu Asp Glu Gly Asp Asp Leu Leu Leu Arg Ser Val Asp Glu Phe
            365                 370                 375
Trp Trp Phe Pro His Met Trp Ser His Met Gln Pro His Leu Phe
            380                 385                 390
His Asn Glu Ser Ser Leu Val Glu Gln Met Ile Leu Asn Lys Glu
```

```
                395                 400                 405
Phe Ala Leu Glu His Gly Ile Pro Ile Asn Met Gly Tyr Ala Val
                410                 415                 420
Ala Pro His His Ser Gly Val Tyr Pro Val His Ile Gln Leu Tyr
                425                 430                 435
Ala Ala Trp Lys Lys Val Trp Gly Ile Gln Val Thr Ser Thr Glu
                440                 445                 450
Glu Tyr Pro His Leu Lys Pro Ala Arg Tyr Arg Lys Gly Phe Ile
                455                 460                 465
His Asn Ser Ile Met Val Leu Pro Arg Gln Thr Cys Gly Leu Phe
                470                 475                 480
Thr His Thr Ile Phe Tyr Lys Glu Tyr Pro Gly Gly Pro Gln Glu
                485                 490                 495
Leu Asp Lys Ser Ile Arg Gly Gly Glu Leu Phe Leu Thr Ile Leu
                500                 505                 510
Leu Asn Pro Ile Ser Ile Phe Met Thr His Leu Ser Asn Tyr Gly
                515                 520                 525
Asn Asp Arg Leu Gly Leu Tyr Thr Phe Val Asn Leu Val Asn Phe
                530                 535                 540
Val Gln Ser Trp Thr Asn Leu Lys Leu Gln Thr Leu Pro Pro Val
                545                 550                 555
Gln Leu Ala His Gln Tyr Phe Glu Leu Phe Pro Glu Gln Lys Asp
                560                 565                 570
Pro Leu Trp Gln Asn Pro Cys Asp Asp Lys Arg His Lys Asp Ile
                575                 580                 585
Trp Ser Arg Glu Lys Thr Cys Asp His Leu Pro Lys Phe Leu Val
                590                 595                 600
Ile Gly Pro Gln Lys Thr Gly Thr Thr Ala Leu Tyr Leu Phe Leu
                605                 610                 615
Leu Met His Pro Ser Ile Ile Ser Asn Leu Pro Ser Pro Lys Thr
                620                 625                 630
Phe Glu Glu Val Gln Phe Phe Asn Gly Asn Asn Tyr His Lys Gly
                635                 640                 645
Ile Asp Trp Tyr Met Asp Phe Phe Pro Ile Pro Ser Asn Thr Thr
                650                 655                 660
Ser Asp Phe Tyr Phe Glu Lys Ser Ala Asn Tyr Phe Asp Ser Glu
                665                 670                 675
Val Ala Pro Arg Arg Ala Ala Ala Leu Leu Pro Lys Ala Lys Val
                680                 685                 690
Leu Thr Ile Leu Ile Asn Pro Ala Asp Arg Ala Tyr Ser Trp Tyr
                695                 700                 705
Gln His Gln Arg Ala His Asp Asp Pro Val Ala Leu Lys Tyr Thr
                710                 715                 720
Phe His Glu Val Ile Thr Ala Gly Ser Asp Ala Ser Ser Lys Leu
                725                 730                 735
Arg Ala Leu Gln Asn Arg Cys Leu Val Pro Gly Trp Tyr Ala Thr
                740                 745                 750
His Ile Glu Arg Trp Leu Ser Ala Tyr His Ala Asn Gln Val Leu
                755                 760                 765
Val Leu Asp Gly Lys Leu Leu Arg Thr Glu Pro Ala Lys Val Met
                770                 775                 780
Asp Met Val Gln Lys Phe Leu Gly Val Thr Asn Thr Ile Asp Tyr
                785                 790                 795
His Lys Thr Leu Ala Phe Asp Pro Lys Lys Gly Phe Trp Cys Gln
                800                 805                 810
Leu Leu Glu Gly Gly Lys Thr Lys Cys Leu Gly Lys Ser Lys Gly
```

```
                815                 820                 825
Arg Lys Tyr Pro Glu Met Asp Leu Gln Ser Arg Ala Phe Leu Lys
                830                 835                 840
Asp Tyr Tyr Arg Asp His Asn Ile Glu Leu Ser Lys Leu Leu Tyr
                845                 850                 855
Lys Met Gly Gln Thr Leu Pro Thr Trp Leu Arg Glu Asp Phe Gln
                860                 865                 870
Asn Thr Arg
```

<210> 4
<211> 581
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2860635CD1

<400> 4

```
Met Trp Gly Arg Thr Ala Arg Arg Arg Cys Pro Arg Glu Leu Arg
  1               5                  10                  15
Arg Gly Arg Glu Ala Leu Leu Val Leu Leu Ala Leu Leu Ala Leu
                20                  25                  30
Ala Gly Leu Gly Ser Val Leu Arg Ala Gln Arg Gly Ala Gly Ala
                35                  40                  45
Gly Ala Ala Glu Pro Gly Pro Pro Arg Thr Pro Arg Pro Gly Arg
                50                  55                  60
Arg Glu Pro Val Met Pro Arg Pro Pro Val Pro Ala Asn Ala Leu
                65                  70                  75
Gly Ala Arg Gly Glu Ala Val Arg Leu Gln Leu Gln Gly Glu Glu
                80                  85                  90
Leu Arg Leu Gln Glu Glu Ser Val Arg Leu His Gln Ile Asn Ile
                95                 100                 105
Tyr Leu Ser Asp Arg Ile Ser Leu His Arg Arg Leu Pro Glu Arg
               110                 115                 120
Trp Asn Pro Leu Cys Lys Glu Lys Lys Tyr Asp Tyr Asp Asn Leu
               125                 130                 135
Pro Arg Thr Ser Val Ile Ile Ala Phe Tyr Asn Glu Ala Trp Ser
               140                 145                 150
Thr Leu Leu Arg Thr Val Tyr Ser Val Leu Glu Thr Ser Pro Asp
               155                 160                 165
Ile Leu Leu Glu Glu Val Ile Leu Val Asp Asp Tyr Ser Asp Arg
               170                 175                 180
Glu His Leu Lys Glu Arg Leu Ala Asn Glu Leu Ser Gly Leu Pro
               185                 190                 195
Lys Val Arg Leu Ile Arg Ala Asn Lys Arg Glu Gly Leu Val Arg
               200                 205                 210
Ala Arg Leu Leu Gly Ala Ser Ala Ala Arg Gly Asp Val Leu Thr
               215                 220                 225
Phe Leu Asp Cys His Cys Glu Cys His Glu Gly Trp Leu Glu Pro
               230                 235                 240
Leu Leu Gln Arg Ile His Glu Glu Glu Ser Ala Val Val Cys Pro
               245                 250                 255
Val Ile Asp Val Ile Asp Trp Asn Thr Phe Glu Tyr Leu Gly Asn
               260                 265                 270
```

```
Ser Gly Glu Pro Gln Ile Gly Gly Phe Asp Trp Arg Leu Val Phe
                275                 280                 285
Thr Trp His Thr Val Pro Glu Arg Glu Arg Ile Arg Met Gln Ser
                290                 295                 300
Pro Val Asp Val Ile Arg Ser Pro Thr Met Ala Gly Gly Leu Phe
                305                 310                 315
Ala Val Ser Lys Lys Tyr Phe Glu Tyr Leu Gly Ser Tyr Asp Thr
                320                 325                 330
Gly Met Glu Val Trp Gly Gly Glu Asn Leu Glu Phe Ser Phe Arg
                335                 340                 345
Ile Trp Gln Cys Gly Gly Val Leu Glu Thr His Pro Cys Ser His
                350                 355                 360
Val Gly His Val Phe Pro Lys Gln Ala Pro Tyr Ser Arg Asn Lys
                365                 370                 375
Ala Leu Ala Asn Ser Val Arg Ala Ala Glu Val Trp Met Asp Glu
                380                 385                 390
Phe Lys Glu Leu Tyr Tyr His Arg Asn Pro Arg Ala Arg Leu Glu
                395                 400                 405
Pro Phe Gly Asp Val Thr Glu Arg Lys Gln Leu Arg Asp Lys Leu
                410                 415                 420
Gln Cys Lys Asp Phe Lys Trp Phe Leu Glu Thr Val Tyr Pro Glu
                425                 430                 435
Leu His Val Pro Glu Asp Arg Pro Gly Phe Phe Gly Met Leu Gln
                440                 445                 450
Asn Lys Gly Leu Thr Asp Tyr Cys Phe Asp Tyr Asn Pro Pro Asp
                455                 460                 465
Glu Asn Gln Ile Val Gly His Gln Val Ile Leu Tyr Leu Cys His
                470                 475                 480
Gly Met Gly Gln Asn Gln Phe Phe Glu Tyr Thr Ser Gln Lys Glu
                485                 490                 495
Ile Arg Tyr Asn Thr His Gln Pro Glu Gly Cys Ile Ala Val Glu
                500                 505                 510
Ala Gly Met Asp Thr Leu Ile Met His Leu Cys Glu Glu Thr Ala
                515                 520                 525
Pro Glu Asn Gln Lys Phe Ile Leu Gln Glu Asp Gly Ser Leu Phe
                530                 535                 540
His Glu Gln Ser Lys Lys Cys Val Gln Ala Ala Arg Lys Glu Ser
                545                 550                 555
Ser Asp Ser Phe Val Pro Leu Leu Arg Asp Cys Thr Asn Ser Asp
                560                 565                 570
His Gln Lys Trp Phe Phe Lys Glu Arg Met Leu
                575                 580
```

<210> 5
<211> 835
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2530615CD1

<400> 5

```
Met Trp Leu Leu Ala Leu Cys Leu Val Gly Leu Ala Gly Ala Gln
  1               5                  10                  15
Arg Gly Gly Gly Gly Pro Gly Gly Gly Ala Pro Gly Gly Pro Gly
```

```
                  20                  25                  30
Leu Gly Leu Gly Ser Leu Gly Glu Glu Arg Phe Pro Val Val Asn
                  35                  40                  45
Thr Ala Tyr Gly Arg Val Arg Gly Val Arg Arg Glu Leu Asn Asn
                  50                  55                  60
Glu Ile Leu Gly Pro Val Val Gln Phe Leu Gly Val Pro Tyr Ala
                  65                  70                  75
Thr Pro Pro Leu Gly Ala Arg Arg Phe Gln Pro Pro Glu Ala Pro
                  80                  85                  90
Ala Ser Trp Pro Gly Val Arg Asn Ala Thr Thr Leu Pro Pro Ala
                  95                 100                 105
Cys Pro Gln Asn Leu His Gly Ala Leu Pro Ala Ile Met Leu Pro
                 110                 115                 120
Val Trp Phe Thr Asp Asn Leu Glu Ala Ala Ala Thr Tyr Val Gln
                 125                 130                 135
Asn Gln Ser Glu Asp Cys Leu Tyr Leu Asn Leu Tyr Val Pro Thr
                 140                 145                 150
Glu Asp Gly Pro Leu Thr Lys Lys Arg Asp Glu Ala Thr Leu Asn
                 155                 160                 165
Pro Pro Asp Thr Asp Ile Arg Asp Pro Gly Lys Lys Pro Val Met
                 170                 175                 180
Leu Phe Leu His Gly Gly Ser Tyr Met Glu Gly Thr Gly Asn Met
                 185                 190                 195
Phe Asp Gly Ser Val Leu Ala Ala Tyr Gly Asn Val Ile Val Ala
                 200                 205                 210
Thr Leu Asn Tyr Arg Leu Gly Val Leu Gly Phe Leu Ser Thr Gly
                 215                 220                 225
Asp Gln Ala Ala Lys Gly Asn Tyr Gly Leu Leu Asp Gln Ile Gln
                 230                 235                 240
Ala Leu Arg Trp Leu Ser Glu Asn Ile Ala His Phe Gly Gly Asp
                 245                 250                 255
Pro Glu Arg Ile Thr Ile Phe Gly Ser Gly Ala Gly Ala Ser Cys
                 260                 265                 270
Val Asn Leu Leu Ile Leu Ser His His Ser Glu Gly Leu Phe Gln
                 275                 280                 285
Lys Ala Ile Ala Gln Ser Gly Thr Ala Ile Ser Ser Trp Ser Val
                 290                 295                 300
Asn Tyr Gln Pro Leu Lys Tyr Thr Arg Leu Leu Ala Ala Lys Val
                 305                 310                 315
Gly Cys Asp Arg Glu Asp Ser Ala Glu Ala Val Glu Cys Leu Arg
                 320                 325                 330
Arg Lys Pro Ser Arg Glu Leu Val Asp Gln Asp Val Gln Pro Ala
                 335                 340                 345
Arg Tyr His Ile Ala Phe Gly Pro Val Val Asp Gly Asp Val Val
                 350                 355                 360
Pro Asp Asp Pro Glu Ile Leu Met Gln Gln Gly Glu Phe Leu Asn
                 365                 370                 375
Tyr Asp Met Leu Ile Gly Val Asn Gln Gly Glu Gly Leu Lys Phe
                 380                 385                 390
Val Glu Asp Ser Ala Glu Ser Glu Asp Gly Val Ser Ala Ser Ala
                 395                 400                 405
Phe Asp Phe Thr Val Ser Asn Phe Val Asp Asn Leu Tyr Gly Tyr
                 410                 415                 420
Pro Glu Gly Lys Asp Val Leu Arg Glu Thr Ile Lys Phe Met Tyr
                 425                 430                 435
Thr Asp Trp Ala Asp Arg Asp Asn Gly Glu Met Arg Arg Lys Thr
```

```
                440                 445                 450
Leu Leu Ala Leu Phe Thr Asp His Gln Trp Val Ala Pro Ala Val
                455                 460                 465
Ala Thr Ala Lys Leu His Ala Asp Tyr Gln Ser Pro Val Tyr Phe
                470                 475                 480
Tyr Thr Phe Tyr His His Cys Gln Ala Glu Gly Arg Pro Glu Trp
                485                 490                 495
Ala Asp Ala Ala His Gly Asp Glu Leu Pro Tyr Val Phe Gly Val
                500                 505                 510
Pro Met Val Gly Ala Thr Asp Leu Phe Pro Cys Asn Phe Ser Lys
                515                 520                 525
Asn Asp Val Met Leu Ser Ala Val Val Met Thr Tyr Trp Thr Asn
                530                 535                 540
Phe Ala Lys Thr Gly Asp Pro Asn Gln Pro Val Pro Gln Asp Thr
                545                 550                 555
Lys Phe Ile His Thr Lys Pro Asn Arg Phe Glu Glu Val Val Trp
                560                 565                 570
Ser Lys Phe Asn Ser Lys Glu Lys Gln Tyr Leu His Ile Gly Leu
                575                 580                 585
Lys Pro Arg Val Arg Asp Asn Tyr Arg Ala Asn Lys Val Ala Phe
                590                 595                 600
Trp Leu Glu Leu Val Pro His Leu His Asn Leu His Thr Glu Leu
                605                 610                 615
Phe Thr Thr Thr Thr Arg Leu Pro Pro Tyr Ala Thr Arg Trp Pro
                620                 625                 630
Pro Arg Pro Pro Ala Gly Ala Pro Gly Thr Arg Arg Pro Pro Pro
                635                 640                 645
Pro Ala Thr Leu Pro Pro Glu Pro Glu Pro Glu Pro Gly Pro Arg
                650                 655                 660
Ala Tyr Asp Arg Phe Pro Gly Asp Ser Arg Asp Tyr Ser Thr Glu
                665                 670                 675
Leu Ser Val Thr Val Ala Val Gly Ala Ser Leu Leu Phe Leu Asn
                680                 685                 690
Ile Leu Ala Phe Ala Ala Leu Tyr Tyr Lys Arg Asp Arg Arg Gln
                695                 700                 705
Glu Leu Arg Cys Arg Arg Leu Ser Pro Pro Gly Gly Ser Gly Ser
                710                 715                 720
Gly Val Pro Gly Gly Gly Pro Leu Leu Pro Ala Ala Gly Arg Glu
                725                 730                 735
Leu Pro Pro Glu Glu Glu Leu Val Ser Leu Gln Leu Lys Arg Gly
                740                 745                 750
Gly Gly Val Gly Ala Asp Pro Ala Glu Ala Leu Arg Pro Ala Cys
                755                 760                 765
Pro Pro Asp Tyr Thr Leu Ala Leu Arg Arg Ala Pro Asp Asp Val
                770                 775                 780
Pro Leu Leu Ala Pro Gly Ala Leu Thr Leu Leu Pro Ser Gly Leu
                785                 790                 795
Gly Pro Pro Pro Pro Pro Pro Pro Pro Ser Leu His Pro Phe Gly
                800                 805                 810
Pro Phe Pro Pro Pro Pro Pro Thr Ala Thr Ser His Asn Asn Thr
                815                 820                 825
Leu Pro His Pro His Ser Thr Thr Arg Val
                830                 835
```

<210> 6
<211> 558

<212> PRT.
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3883906CD1

<400> 6

```
Met Arg Lys Ile Arg Ala Asn Ala Ile Ala Ile Leu Thr Val Ala
  1               5                  10                  15
Trp Ile Leu Gly Thr Phe Tyr Tyr Leu Trp Gln Asp Asn Arg Ala
                 20                  25                  30
His Ala Ala Ser Ser Gly Gly Arg Gly Ala Gln Arg Ala Gly Arg
                 35                  40                  45
Arg Ser Glu Gln Leu Arg Glu Asp Arg Thr Ile Pro Leu Ile Val
                 50                  55                  60
Thr Gly Thr Pro Ser Lys Gly Phe Asp Glu Lys Ala Tyr Leu Ser
                 65                  70                  75
Ala Lys Gln Leu Lys Ala Gly Glu Asp Pro Tyr Arg Gln His Ala
                 80                  85                  90
Phe Asn Gln Leu Glu Ser Asp Lys Leu Ser Pro Asp Arg Pro Ile
                 95                 100                 105
Arg Asp Thr Arg His Tyr Ser Cys Pro Ser Val Ser Tyr Ser Ser
                110                 115                 120
Asp Leu Pro Ala Thr Ser Val Ile Ile Thr Phe His Asn Glu Ala
                125                 130                 135
Arg Ser Thr Leu Leu Arg Thr Val Lys Ser Val Leu Asn Arg Thr
                140                 145                 150
Pro Ala Asn Leu Ile Gln Glu Ile Ile Leu Val Asp Asp Phe Ser
                155                 160                 165
Ser Asp Pro Glu Asp Cys Leu Leu Leu Thr Arg Ile Pro Lys Val
                170                 175                 180
Lys Cys Leu Arg Asn Asp Arg Arg Glu Gly Leu Ile Arg Ser Arg
                185                 190                 195
Val Arg Gly Ala Asp Val Ala Ala Ala Thr Val Leu Thr Phe Leu
                200                 205                 210
Asp Ser His Cys Glu Val Asn Thr Glu Trp Leu Pro Pro Met Leu
                215                 220                 225
Gln Arg Val Lys Glu Asp His Thr Arg Val Val Ser Pro Ile Ile
                230                 235                 240
Asp Val Ile Ser Leu Asp Asn Phe Ala Tyr Leu Ala Ala Ser Ala
                245                 250                 255
Asp Leu Arg Gly Gly Phe Asp Trp Ser Leu His Phe Lys Trp Glu
                260                 265                 270
Gln Ile Pro Leu Glu Gln Lys Met Thr Arg Thr Asp Pro Thr Arg
                275                 280                 285
Pro Ile Arg Thr Pro Val Ile Ala Gly Gly Ile Phe Val Ile Asp
                290                 295                 300
Lys Ser Trp Phe Asn His Leu Gly Lys Tyr Asp Ala Gln Met Asp
                305                 310                 315
Ile Trp Gly Gly Glu Asn Phe Glu Leu Ser Phe Arg Val Trp Met
                320                 325                 330
Cys Gly Gly Ser Leu Glu Ile Val Pro Cys Ser Arg Val Gly His
                335                 340                 345
Val Phe Arg Lys Arg His Pro Tyr Asn Phe Pro Glu Gly Asn Ala
                350                 355                 360
```

```
Leu Thr Tyr Ile Arg Asn Thr Lys Arg Thr Ala Glu Val Trp Met
                365                 370                 375
Asp Glu Tyr Lys Gln Tyr Tyr Tyr Glu Ala Arg Pro Ser Ala Ile
                380                 385                 390
Gly Lys Ala Phe Gly Ser Val Ala Thr Arg Ile Glu Gln Arg Lys
                395                 400                 405
Lys Met Asn Cys Lys Ser Phe Arg Trp Tyr Leu Glu Asn Val Tyr
                410                 415                 420
Pro Glu Leu Thr Val Pro Val Lys Glu Ala Leu Pro Gly Ile Ile
                425                 430                 435
Lys Gln Gly Val Asn Cys Leu Glu Ser Gln Gly Gln Asn Thr Ala
                440                 445                 450
Gly Asp Phe Leu Leu Gly Met Gly Ile Cys Arg Gly Ser Ala Lys
                455                 460                 465
Asn Pro Gln Pro Ala Gln Ala Trp Leu Phe Ser Asp His Leu Ile
                470                 475                 480
Gln Gln Gln Gly Lys Cys Leu Ala Ala Thr Ser Thr Leu Met Ser
                485                 490                 495
Ser Pro Gly Ser Pro Val Ile Leu Gln Met Cys Asn Pro Arg Glu
                500                 505                 510
Gly Lys Gln Lys Trp Arg Arg Lys Gly Ser Phe Ile Gln His Ser
                515                 520                 525
Val Ser Gly Leu Cys Leu Glu Thr Lys Pro Ala Gln Leu Val Thr
                530                 535                 540
Ser Lys Cys Gln Ala Asp Ala Gln Ala Gln Gln Trp Gln Leu Leu
                545                 550                 555
Pro His Thr
```

<210> 7
<211> 277
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473644CD1

<400> 7

```
Met Ser Ser Cys Ser Arg Val Ala Leu Val Thr Gly Ala Asn Lys
  1               5                  10                  15
Gly Ile Gly Phe Ala Ile Thr Arg Asp Leu Cys Arg Lys Phe Ser
                 20                  25                  30
Gly Asp Val Val Leu Thr Ala Arg Asp Glu Ala Arg Gly Arg Ala
                 35                  40                  45
Ala Val Gln Gln Leu Gln Ala Glu Gly Leu Ser Pro Arg Phe His
                 50                  55                  60
Gln Leu Asp Ile Asp Asp Pro Gln Ser Ile Arg Ala Leu Arg Asp
                 65                  70                  75
Phe Leu Arg Lys Glu Tyr Gly Gly Leu Asn Val Leu Val Asn Asn
                 80                  85                  90
Ala Gly Ile Ala Phe Arg Met Asp Asp Pro Thr Pro Phe Asp Ile
                 95                 100                 105
Gln Ala Glu Val Thr Leu Lys Thr Asn Phe Phe Ala Thr Arg Asn
                110                 115                 120
Val Cys Thr Glu Leu Leu Pro Ile Met Lys Pro His Gly Arg Val
```

```
                125                 130                 135
Val Asn Ile Ser Ser Leu Ile Ser Leu Glu Ala Leu Lys Asn Cys
                140                 145                 150
Ser Leu Glu Leu Gln Gln Lys Phe Arg Ser Glu Thr Ile Thr Glu
                155                 160                 165
Glu Glu Leu Val Gly Leu Met Asn Lys Phe Val Glu Asp Thr Lys
                170                 175                 180
Lys Gly Val His Ala Lys Glu Gly Trp Pro Asn Ser Ala Tyr Gly
                185                 190                 195
Val Ser Lys Ile Gly Val Thr Val Leu Ser Arg Ile Leu Ala Arg
                200                 205                 210
Lys Leu Asn Glu Gln Arg Arg Gly Asp Lys Ile Leu Leu Asn Ala
                215                 220                 225
Cys Cys Pro Gly Trp Val Arg Thr Asp Met Ala Gly Pro Gln Ala
                230                 235                 240
Thr Lys Ser Pro Glu Glu Gly Ala Glu Thr Pro Val Tyr Leu Ala
                245                 250                 255
Leu Leu Pro Pro Asp Ala Glu Gly Pro His Gly Gln Phe Val Gln
                260                 265                 270
Asp Lys Lys Val Glu Gln Trp
                275
```

<210> 8
<211> 821
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485303CD1

<400> 8

```
Met Val Asn Gly Leu Phe Leu Phe Ser Leu Tyr Thr Pro Gln Ala
  1                 5                  10                  15
Pro Ser Met Ser Ser Lys Asn Asn Leu Glu Ala Phe Arg Arg Pro
                 20                  25                  30
Glu Gln Arg Gly His Arg Ala Ala Arg Ser Gln Leu Arg Ser Asp
                 35                  40                  45
Ala Ala Ser Ala Arg Arg Gly Phe Val Ala Arg Thr Arg Arg Glu
                 50                  55                  60
Leu Gly Pro Gly Arg Cys Val Arg Pro Arg Pro Arg Leu Leu Ala
                 65                  70                  75
Pro Thr Ser Leu Asn Asn Asp His Arg Gln Pro Arg Trp Glu Asp
                 80                  85                  90
Asn Gly Gln Lys Met Pro Gly His Pro Glu Thr Leu Ser Ala Phe
                 95                 100                 105
Cys Pro Gly His Arg Leu Leu His Ser Leu Leu Pro Asp Asp Val
                110                 115                 120
Ser Gly Arg Pro Val Leu Gly Pro Ala Pro Gly Gly Leu Ala Ile
                125                 130                 135
His His Pro His Arg Arg Met Pro Pro Gly Ala Gly Leu Met Glu
                140                 145                 150
Arg Ile Gln Ala Ile Ala Gln Asn Val Ser Asp Ile Ala Val Lys
                155                 160                 165
Val Asp Gln Ile Leu Arg His Ser Leu Leu Leu His Ser Lys Val
                170                 175                 180
```

```
Ser Glu Gly Arg Arg Asp Gln Cys Glu Ala Pro Ser Asp Pro Lys
                185                 190                 195
Phe Pro Asp Cys Ser Gly Lys Val Glu Trp Met Arg Ala Arg Trp
                200                 205                 210
Thr Ser Asp Pro Cys Tyr Ala Phe Phe Gly Val Asp Gly Thr Glu
                215                 220                 225
Cys Ser Phe Leu Ile Tyr Leu Ser Glu Val Glu Trp Phe Cys Pro
                230                 235                 240
Pro Leu Pro Trp Arg Asn Gln Thr Ala Ala Gln Arg Ala Pro Lys
                245                 250                 255
Pro Leu Pro Lys Val Gln Ala Val Phe Arg Ser Asn Leu Ser His
                260                 265                 270
Leu Leu Asp Leu Met Gly Ser Gly Lys Glu Ser Leu Ile Phe Met
                275                 280                 285
Lys Lys Arg Thr Lys Arg Leu Thr Ala Gln Trp Ala Leu Ala Ala
                290                 295                 300
Gln Arg Leu Ala Gln Lys Leu Gly Ala Thr Gln Arg Asp Gln Lys
                305                 310                 315
Gln Ile Leu Val His Ile Gly Phe Leu Thr Glu Glu Ser Gly Asp
                320                 325                 330
Val Phe Ser Pro Arg Val Leu Lys Gly Gly Pro Leu Gly Glu Met
                335                 340                 345
Val Gln Trp Ala Asp Ile Leu Thr Ala Leu Tyr Val Leu Gly His
                350                 355                 360
Gly Leu Arg Val Thr Val Ser Leu Lys Glu Leu Gln Ser Asn Leu
                365                 370                 375
Gly Val Pro Pro Gly Arg Gly Ser Cys Pro Leu Thr Met Pro Leu
                380                 385                 390
Pro Phe Asp Leu Ile Tyr Thr Asp Tyr His Gly Leu Gln Gln Met
                395                 400                 405
Lys Arg His Met Gly Leu Ser Phe Lys Lys Tyr Arg Cys Arg Ile
                410                 415                 420
Arg Val Ile Asp Thr Phe Gly Thr Glu Pro Ala Tyr Asn His Glu
                425                 430                 435
Glu Tyr Ala Thr Leu His Gly Tyr Arg Thr Asn Trp Gly Tyr Trp
                440                 445                 450
Asn Leu Asn Pro Lys Gln Phe Met Thr Met Phe Pro His Thr Pro
                455                 460                 465
Asp Asn Ser Phe Met Gly Phe Val Ser Glu Glu Leu Asn Glu Thr
                470                 475                 480
Glu Lys Arg Leu Ile Lys Gly Gly Lys Ala Ser Asn Met Ala Val
                485                 490                 495
Val Tyr Gly Lys Glu Ala Ser Ile Trp Lys Gly Lys Glu Lys Phe
                500                 505                 510
Leu Gly Ile Leu Asn Lys Tyr Met Glu Ile His Gly Thr Val Tyr
                515                 520                 525
Tyr Glu Ser Gln Arg Pro Pro Glu Val Pro Ala Phe Val Lys Asn
                530                 535                 540
His Gly Leu Leu Pro Gln Pro Glu Phe Gln Gln Leu Leu Arg Lys
                545                 550                 555
Ala Lys Leu Phe Ile Gly Phe Gly Phe Pro Tyr Glu Gly Pro Ala
                560                 565                 570
Pro Leu Glu Ala Ile Ala Asn Gly Cys Ile Phe Leu Gln Ser Arg
                575                 580                 585
Phe Ser Pro Pro His Ser Ser Leu Asn His Glu Phe Phe Arg Gly
                590                 595                 600
```

```
Lys Pro Thr Ser Arg Glu Val Phe Ser Gln His Pro Tyr Ala Glu
                605                 610                 615
Asn Phe Ile Gly Lys Pro His Val Trp Thr Val Asp Tyr Asn Asn
                620                 625                 630
Ser Glu Glu Phe Glu Ala Ala Ile Lys Ala Ile Met Arg Thr Gln
                635                 640                 645
Val Asp Pro Tyr Leu Pro Tyr Glu Tyr Thr Cys Glu Gly Met Leu
                650                 655                 660
Glu Arg Ile His Ala Tyr Ile Gln His Gln Asp Phe Cys Arg Ala
                665                 670                 675
Pro Asp Pro Ala Leu Pro Glu Ala His Ala Pro Gln Ser Pro Phe
                680                 685                 690
Val Leu Ala Pro Asn Ala Thr His Leu Glu Trp Ala Arg Asn Thr
                695                 700                 705
Ser Leu Ala Pro Gly Ala Trp Pro Pro Ala His Ala Leu Arg Ala
                710                 715                 720
Trp Leu Ala Val Pro Gly Arg Ala Cys Thr Asp Thr Cys Leu Asp
                725                 730                 735
His Gly Leu Ile Cys Glu Pro Ser Phe Phe Pro Phe Leu Asn Ser
                740                 745                 750
Gln Asp Ala Phe Leu Lys Leu Gln Val Pro Cys Asp Ser Thr Glu
                755                 760                 765
Ser Glu Met Asn His Leu Tyr Pro Ala Phe Ala Gln Pro Gly Gln
                770                 775                 780
Glu Cys Tyr Leu Gln Lys Glu Pro Leu Leu Phe Ser Cys Ala Gly
                785                 790                 795
Ser Asn Thr Lys Tyr Arg Arg Leu Cys Pro Cys Arg Asp Phe Arg
                800                 805                 810
Lys Gly Gln Val Ala Leu Cys Gln Gly Cys Leu
                815                 820
```

<210> 9
<211> 159
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6970969CD1

<400> 9

```
Met Thr Ile Glu Ile Arg Pro Ala Val Pro Ser Asp Ala Ala Gln
  1                 5                  10                  15
Ile Leu Thr Phe Ile Thr Glu Leu Ala Glu Tyr Glu Lys Ala Arg
                 20                  25                  30
His Glu Val Ile Ala Ser Val Val Asp Ile Glu Arg Ser Leu Phe
                 35                  40                  45
Ser Glu Gly Ala Thr Ala His Gly Leu Ile Cys Leu Arg Asp Gly
                 50                  55                  60
Leu Pro Ile Gly Phe Ala Val Phe Phe Phe Ser Tyr Ser Thr Trp
                 65                  70                  75
Leu Gly Ser Asn Cys Leu Tyr Leu Glu Asp Leu Tyr Ile Asn Pro
                 80                  85                  90
Glu Gln Arg Gly Gly Gly Ala Gly Lys Lys Leu Leu Arg His Leu
                 95                 100                 105
Ala Lys Ile Ala Phe Asp Asn Gly Cys Gly Arg Phe Glu Trp Ser
```

```
                110                 115                 120
Val Leu Asp Trp Asn Glu Pro Ala Ile Ala Phe Tyr Lys Ser Ile
                125                 130                 135
Gly Ala Gln Pro Gln Glu Glu Trp Val Arg Tyr Arg Met Glu Gly
                140                 145                 150
Asp Ala Leu Arg Asp Phe Ala Leu Gly
                155
```

<210> 10
<211> 642
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6538080CD1

<400> 10

```
Met Val Val Val Pro Phe Thr Ile Pro Phe Asp Ser Ser Val Trp
  1                 5                  10                  15
Leu Leu Gln Lys Leu Asn Ser Pro Trp Arg Met Thr Val Asp Tyr
                 20                  25                  30
His Thr Leu Asn Gln Ala Val Thr Leu Ile Ile Ala Ala Val Leu
                 35                  40                  45
Gly Val Ala Ser Leu Leu Glu Gln Gly Phe Ser Arg Ala Ile Trp
                 50                  55                  60
Glu Asn Arg Glu Ser Leu Ser Lys Val Cys Val Ser Gln Gly Pro
                 65                  70                  75
Ser Arg Pro Leu Ala Cys Ala Thr Asn Gly Asp Ile Lys Val Gln
                 80                  85                  90
Gly Gly Pro Ser Ala Glu Gly Pro Gln Arg Asn Thr Arg Leu Gly
                 95                 100                 105
Trp Ile Gln Gly Lys Gln Val Thr Val Leu Gly Ser Pro Val Pro
                110                 115                 120
Val Asn Val Phe Leu Gly Val Pro Phe Ala Ala Pro Pro Leu Gly
                125                 130                 135
Ser Leu Arg Phe Thr Asn Pro Gln Pro Ala Ser Pro Trp Asp Asn
                140                 145                 150
Leu Arg Glu Ala Thr Ser Tyr Pro Asn Leu Cys Leu Gln Asn Ser
                155                 160                 165
Glu Trp Leu Leu Leu Asp Gln His Met Leu Lys Val His Tyr Pro
                170                 175                 180
Lys Phe Gly Val Ser Glu Asp Cys Leu Tyr Leu Asn Ile Tyr Ala
                185                 190                 195
Pro Ala His Ala Asp Thr Gly Ser Lys Leu Pro Val Leu Val Trp
                200                 205                 210
Phe Pro Gly Gly Ala Phe Lys Thr Gly Ser Ala Ser Ile Phe Asp
                215                 220                 225
Gly Ser Ala Leu Ala Ala Tyr Glu Asp Val Leu Val Val Val Val
                230                 235                 240
Gln Tyr Arg Leu Gly Ile Phe Gly Phe Phe Thr Thr Trp Asp Gln
                245                 250                 255
His Ala Pro Gly Asn Trp Ala Phe Lys Asp Gln Val Ala Ala Leu
                260                 265                 270
Ser Trp Val Gln Lys Asn Ile Glu Phe Phe Gly Gly Asp Pro Ser
                275                 280                 285
```

```
Ser Val Thr Ile Phe Gly Glu Ser Ala Gly Ala Ile Ser Val Ser
                290                 295                 300
Ser Leu Ile Leu Ser Pro Met Ala Lys Gly Leu Phe His Lys Ala
                305                 310                 315
Ile Met Glu Ser Gly Val Ala Ile Ile Pro Tyr Leu Glu Ala His
                320                 325                 330
Asp Tyr Glu Lys Ser Glu Asp Leu Gln Val Val Ala His Phe Cys
                335                 340                 345
Gly Asn Asn Ala Ser Asp Ser Glu Ala Leu Leu Arg Cys Leu Arg
                350                 355                 360
Thr Lys Pro Ser Lys Glu Leu Leu Thr Leu Ser Gln Lys Thr Lys
                365                 370                 375
Ser Phe Thr Arg Val Val Asp Gly Ala Phe Phe Pro Asn Glu Pro
                380                 385                 390
Leu Asp Leu Leu Ser Gln Lys Ala Phe Lys Ala Ile Pro Ser Ile
                395                 400                 405
Ile Gly Val Asn Asn His Glu Cys Gly Phe Leu Leu Pro Met Lys
                410                 415                 420
Glu Ala Pro Glu Val Leu Ser Gly Ser Asn Lys Ser Leu Ala Leu
                425                 430                 435
His Leu Ile Gln Asn Ile Leu His Ile Pro Pro Gln Tyr Leu His
                440                 445                 450
Leu Val Ala Asn Glu Tyr Phe His Asp Lys His Ser Leu Thr Glu
                455                 460                 465
Ile Arg Asp Ser Leu Leu Asp Leu Leu Gly Asp Val Phe Phe Val
                470                 475                 480
Val Pro Ala Leu Ile Thr Ala Arg Tyr His Arg Asp Ala Gly Ala
                485                 490                 495
Pro Val Tyr Phe Tyr Glu Phe Arg His Arg Pro Gln Cys Phe Glu
                500                 505                 510
Asp Thr Lys Pro Ala Phe Val Lys Ala Asp His Ala Asp Glu Val
                515                 520                 525
Arg Phe Val Phe Gly Gly Ala Phe Leu Lys Gly Asp Ile Val Met
                530                 535                 540
Phe Glu Gly Ala Thr Glu Glu Glu Lys Leu Leu Ser Arg Lys Met
                545                 550                 555
Met Lys Tyr Trp Ala Thr Phe Ala Arg Thr Gly Asn Pro Asn Gly
                560                 565                 570
Asn Asp Leu Phe Leu Trp Pro Ala Tyr Asn Leu Thr Glu Gln Tyr
                575                 580                 585
Leu Gln Leu Asp Leu Asn Met Ser Leu Gly Gln Arg Leu Lys Glu
                590                 595                 600
Pro Arg Val Glu Phe Trp Thr Ser Thr Ile Pro Leu Ile Leu Ser
                605                 610                 615
Ala Ser Asp Met Leu His Ser Pro Leu Ser Ser Leu Thr Phe Leu
                620                 625                 630
Ser Leu Leu Gln Pro Phe Phe Phe Phe Cys Ala Pro
                635                 640

<210> 11
<211> 109
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
```

<223> Incyte ID No: 55048919CD1

<400> 11
Met His Phe His Ala Thr Gly Asp Cys Ser Asp Gln Gly Phe Gln
1 5 10 15
Lys Ser Gly Ala His Ile Asn His Glu Asp His Lys Thr Pro His
20 25 30
Gly Leu Leu Asn Pro Glu Gly Pro Asp Phe Gly Glu Leu Pro Asn
35 40 45
Ile His Val Ala Ala Asp Gly Thr Val Asn Ala Glu Ala Phe Ser
50 55 60
Ala Leu Val Ser Leu Asp Ala Ala Ser Ser Arg Pro Asn Leu Leu
65 70 75
Asp Ala Asp Gly Ser Ala Leu Val Ile His Ala Ser Pro Asp Asp
80 85 90
His Val Thr Gln Pro Ile Gly Gly Ala Gly Ala Arg Val Ala Cys
95 100 105
Ala Val Ile Arg

<210> 12
<211> 222
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485135CD1

<400> 12
Met Ala Glu Lys Pro Lys Leu His Tyr Ser Asn Ala Gln Gly Arg
1 5 10 15
Met Glu Ser Thr Trp Trp Leu Leu Ala Ala Ala Gly Val Glu Phe
20 25 30
Glu Glu Lys Phe Leu Glu Ser Ala Glu Asp Leu Glu Lys Leu Arg
35 40 45
Asn Asp Gly Ser Phe Met Phe Gln Gln Val Pro Met Val Glu Ile
50 55 60
Asp Gly Met Lys Leu Val Gln Thr Arg Ala Ile Leu Asn Tyr Val
65 70 75
Ala Ser Lys Tyr Asn Leu Tyr Arg Lys Asp Ile Lys Glu Arg Val
80 85 90
Leu Ile Asp Met Tyr Thr Glu Gly Ile Ala Asp Leu Gly Glu Met
95 100 105
Ile Leu Pro Leu Pro Ile Cys Gln Pro Glu Asp Lys Asp Thr Lys
110 115 120
Leu Ala Leu Asn Gln Gln Arg Thr Lys Asn Arg Tyr Phe Pro Ala
125 130 135
Phe Glu Lys Val Leu Lys Ser Asn Gly Gln Asp Tyr Leu Val Gly
140 145 150
Asn Lys Leu Ser Trp Ala Asp Ile His Leu Val Glu Leu Leu Tyr
155 160 165
Tyr Met Glu Glu Leu Asp Ser Ser Leu Ile Phe Ser Phe Pro Leu
170 175 180
Leu Lys Ala Leu Lys Thr Arg Ile Ser Asn Leu Pro Met Val Lys
185 190 195

```
Lys Phe Leu Gln Pro Gly Ser Pro Arg Lys Ser Leu Met Asp Glu
                200                 205                 210
Lys Ser Leu Glu Glu Ala Arg Lys Ile Phe Arg Phe
                215                 220
```

<210> 13
<211> 273
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7684978CD1

<400> 13

```
Met Arg Tyr Val Lys Leu Ala Gly Cys Ser Val Pro Ala Ile Gly
  1               5                  10                  15
Gln Gly Thr Arg Tyr Met Gly Glu Asp Pro Ala Arg Lys Ala Ala
                20                  25                  30
Glu Val Ala Ala Leu Gln Tyr Gly Ile Glu Gln Gly Met Thr Leu
                35                  40                  45
Ile Asp Ser Ala Glu Met Tyr Ala Glu Gly Gly Ala Glu Lys Val
                50                  55                  60
Val Gly Gln Ala Ile Ala Gly Arg Arg Asp Gln Val Phe Leu Val
                65                  70                  75
Ser Lys Val Tyr Pro His Asn Ala Ser Arg Arg Gly Val Pro Glu
                80                  85                  90
Ala Cys Glu Arg Ser Leu Lys Arg Leu Gly Thr Glu Val Ile Asp
                95                 100                 105
Leu Tyr Leu Leu His Trp Arg Gly Gln Tyr Pro Leu Glu Glu Thr
               110                 115                 120
Val Glu Ala Phe Glu Arg Leu Arg Glu Gln Gly Lys Ile Arg His
               125                 130                 135
Trp Gly Val Ser Asn Phe Asp Val Asp Asp Leu Arg Glu Leu His
               140                 145                 150
Asn Pro Asp Cys Ala Thr Asn Gln Val Leu Tyr Asn Pro Ala Gln
               155                 160                 165
Arg Gly Ile Glu Phe Asp Leu Leu Pro Trp Ser Arg Gln Arg Gly
               170                 175                 180
Leu Pro Thr Met Ala Tyr Cys Pro Leu Ala Gln Ala Gly Gln Leu
               185                 190                 195
Leu Gln His Pro Val Leu Ser Glu Ile Ala Glu Arg His Gly Ala
               200                 205                 210
Arg Pro Ala Gln Val Ser Leu Ala Trp Val Thr Arg Asp Asn Asp
               215                 220                 225
Val Ile Ala Ile Pro Lys Ala Val Ser Ser Glu His Val Arg Leu
               230                 235                 240
Asn Ala Ala Ala Gly Ala Leu Thr Leu Thr Ser Glu Asp Leu Arg
               245                 250                 255
Ala Ile Asp Arg Ala Phe Pro Ala Pro Thr Arg Lys Gln His Leu
               260                 265                 270
Ala Met Val
```

<210> 14
<211> 2004

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7483477CB1

<400> 14
catacactcc acctggggac agaaggaaaa tctggtctca aaaaataaaa cattaacaaa 60
aataaaacaa catgttttcc ctatctgaat agatgggggа aacaagtccc cagaggttaa 120
gtgactttcc cagtgaagaa aagacagagc tgccacttgt acccagattg tctggatcct 180
aaaatcaata gcctaaaaag ataataattt gcctcttaca ccccaccaag aaattctgtt 240
caattctgtt gagctgagag agaagaacag ggagactgga aatatgaaga aaggggaatg 300
gccccagagg gtttccctgt gtgaattcaa tttaccttta ttcaggtggg agattatagg 360
ttatataacc acccccaata taaagcaatc tcttccccga tgacactcaa acccggaaaa 420
cttgaatccc agggtgttca gattttggta taagctgatc tcccaggacc tacttaggca 480
ggatctggtg cccaggacag aacctgttga cctggcacca tggagctggg aggggctgtc 540
accatctttc tggcactccg cttgtcttgc ctgctcatcc tcattgcctg gaaacgaatg 600
gataaggcag gaaagctgcc cccaggtcct acaccaattc ctttcctggg gaacctgctg 660
caagttcgaa ctgacgctac gtttcagtct ttcatgaagc tcagggagaa atacagccct 720
gtgttcactg tgtacatggg tccccggcca gtggtggttt tatgtggaca cgaagcagtg 780
aaggaagcct tgatagacca agcggatgag ttcagtggcc gtggagaact ggcttcaata 840
aagcaaaact tccaaggtca tggtgtagct ctggccaatg gagaacgatg gaggattctc 900
cgacgcttct ccctgaccat ccttcgggac ttcgggatgg gaaagcagag cattaaggag 960
cgaattcagg aggaggccag ctacctactg gaggagttcc agaagaccaa gggtgccccc 1020
atcgacccca tcttcttgct gagccgcacc gtttccaacg tcatcagttc cgttgtcttc 1080
agaagtcgct tcgactatga ggacaagcag ttcctgaatc tgctgcggtt gatcaatgag 1140
agtttcattg agatgagcac tccctgggca cagctatatg acatgtactc tggaatcatg 1200
cagtatttgc caggaagaca caatctcatc tactacttgg tagaagagct caaggacttc 1260
attgcctcca gggtcaagat caacgaagca tcctttgacc cccaaaaccc tcgggacttc 1320
attgactgct ttctcatcaa gatgcaccag gataaaaata atcctcgtac agaattcaac 1380
ctcaagaact tggtcctcac cactctgaac ctcttctttg ctggcacgga gacggtgagc 1440
tctacactgc gctatggatt cttgctgtta atgaagcatc ctgaagtgga agctaagatc 1500
catgaagaga ttaaccaagt gattggacca caccggctcc caagggtgga tgaccgggtc 1560
aagatgccct acacagatgc cgtcatccac gaaatacaga gactggtaga catcgtgccc 1620
atgggtgtcc cccacaacct catccgggac actcagtttc gaggctacct tctgcccaag 1680
ggcacagatg tatttcccct gcttggctct gtcctcaaag accccaaata cttccgctac 1740
ccagatgcct tctatcccca gcactttctg gatgagcagg gccgcttcaa gaagaatgaa 1800
gcctttgtgc ctttttcctc tggaaagcgc atctgcctgg gcgaggccat ggaccgcatg 1860
gaactctttc tctacttcac ctccaccctt cagaacttct ctctacactc gctggtgccg 1920
cctgtggaca tcgatatcac ccccaagctc tcaggctttg gcaacatccc cccgacctac 1980
gagctctgtc ttgtggcccg ctga 2004

<210> 15
<211> 1191
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485159CB1

<400> 15
aaagccccgg ctgccggcgg cggaccacct ctgctgccgc gcgcctaccg gagccgcttg 60
gccctagtgc tttccagcgg atttcccctc agctcctact ctcgggcttc caaatctggg 120
gcgatgtctc cccaggttaa attaccctag ctcctgctcc agatcgcttc cccgtgcccc 180

gccagagccc agtagttcaa aaattaaatt tggggcaagg ggtgcgcgcc agagcgcagc 240
tgtttctgga gcctgcggca gcggtggcga gccacagggc ggcgaccgtg agctccggga 300
gctgcgcaaa ccacctggag accatgtctg gggatgcgac caggaccctg gggaaaggaa 360
gccagccccc agggccagtc ccggaggggc tgatccgcat ctacagcatg aggttctgcc 420
cctattctca caggacccgc ctcgtcctca aggccaaaga catcagacat gaagtggtca 480
acattaacct gagaaacaag cctgaatggt actatacaaa gcacccttt ggccacattc 540
ctgtcctgga gaccagccaa tgtcaactga tctatgaatc tgttattgct tgtgagtacc 600
tggatgatgc ttatccagga aggaagctgt ttccatatga cccttatgaa cgagctcgcc 660
aaaagatgtt attggagcta ttttgtaagg tcccacattt gaccaaggag tgcctggtag 720
cgttgagatg tgggagagaa tgcactaatc tgaaggcagc cctgcgtcag gaattcagca 780
acctggaaga gattcttgag tatcagaaca ccaccttctt tggtggaacc tgtatatcca 840
tgattgatta cctcctctgg ccctggtttg agcggctgga tgtgtatggg atactggact 900
gtgtgagcca cacgccagcc ctgcggctct ggatatcagc catgaagtgg gaccccacag 960
tctgtgctct tctcatggat aagagcattt tccagggctt cttgaatctc tattttcaga 1020
acaaccctaa tgcctttgac tttgggctgt gctgagtctc actgtccacc ccttcgctgt 1080
ccagaattcc ccagcttgtt gggagtctac gtcacggctt gtcttgggaa ccaatccgtc 1140
tctctttctt ttctttgaag ttcccaataa aatgaaaaca ggaaaaaaaa a 1191

<210> 16
<211> 3419
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485518CB1

<400> 16
tttgggcgta tccgttcccc ttacccctttt tacccatgga gccttaactc gcgcccggtg 60
gatcccaatc cctattggga attagagtta gactcccatc aggctactta ctcgattcgg 120
ctccagaaag cccgatacac taacaagagg cacaccacca ctaccacagc agacaggacc 180
actggtattc cgggtgatta aatagttctc atacctttg ctgtactcgt ctgctccctg 240
gtgtaattct cgtgtccaag aaggaactgc attttaacgg tgtgctgaac agaggtttga 300
atcattccag ctgctgcttg ttttgaaaag gacatgtgtg tataaataaa aggactctgc 360
tggaagagtt cagacactaa gactggagct agagagtgct cttggaatat tgttgaggca 420
gattgcatga actgaaagct ccttctaatt aacctggagc caagtgaacc tgaatactgg 480
atatctcatg ttctaacacg ggataaattc aagttagaaa aagacaaata ttgaaatgct 540
tctctagatt atgcttagag gattctgcag cattctcagt attggacata ttcaaatctc 600
ctcaacgtat atccttccca tctattgcct tccagcttta actcagttgt gctcccttca 660
aacatttttac atgatttgcg atggttacat gaactctggg gaagtgatat ttgcaacatc 720
tcatacttta ttcactatgg cactttagga aacgaaattg ggaaaaagct tccaaaacat 780
tattctaaaa aatgaatctt attgtgaaac ttcggagaag ttttcgaaca ttgattgttc 840
tcttagctac cttttgcttg gtgagcattg tcatttctgc ctattttctc tactctggct 900
acaaacagga aatgacactt attgaaacca ctgcagaagc agaatgcact gacatcaaaa 960
ttctaccata taggtcaatg gagctgaaaa cagttaaacc tattgacaca tccaaaacgg 1020
accctactgt ccttctcttc gtggagagcc aatactctca actcggtcaa gatatcatag 1080
ctattttgga gtccagccga tttcagtacc acatggttat tgcccctgga aagggagata 1140
tacctcctct tacagataat ggcaaaggga aatatacttt agttatttat gaaaatattc 1200
tgaagtatgt cagcatggac tcatggaatc gagagctttt agaaaaatac tgtgtggaat 1260
acagtgttag tataatcggt tttcataaag ccaatgagaa cagcttacca agtacacaat 1320
taaaaggctt tccgttaaac cttttcaata atctagctct aaaggactgt tttgttaacc 1380
ctcaatctcc tttgctgcat attaccaaag cccccaaggt tgagaaaggc cctcttcctg 1440
gggaagactg gactattttc caatataatc attcaaccta ccagcctgta cttttaactg 1500
agttacagac agaaaaatcc ctgtcatcct tgtctagcaa aacactcttt gcaacggtga 1560
ttcaggatct ggggcttcat gatggaattc agagagtact ttttggcaac aacttgaact 1620
tttggctgca caagctcatc ttcatagatg ccatctcctt cttgtcaggg aagaggctga 1680

cattgtcctt ggacaggtac atccttgtgg acattgatga tatatttgtt gggaaagagg 1740
gaacaaggat gaatgtcaaa gatgtgaagg taagcttatt tgagactcaa aatttactgc 1800
gcactcaggt tgcaaatttt accttcaacc ttggattttc agggaagttt taccacacag 1860
ggactgaaga ggaagatgaa ggagatgacc ttttacttcg gtctgtggat gagttctggt 1920
ggtttcctca catgtggagc cacatgcagc cccacctctt ccacaacgag tcatctttag 1980
tagagcagat gattctcaac aaggaatttg cactggaaca tggaatacca atcaacatgg 2040
gctatgctgt ggccccacat cactcagggg tctacccggt tcacattcag ctgtatgcag 2100
cttggaagaa ggtctggggt attcaagtca ccagcactga agaatatcca catctgaaac 2160
ctgcccggta cagaaagggc ttcattcaca atagcatcat ggtcctccct cgacagactt 2220
gtgggttgtt cactcacact attttctaca aagaatatcc aggaggaccc caagaactgg 2280
ataaaagtat cagaggaggt gaacttttte tcacaatcct tctaaaccca atcagcattt 2340
tcatgaccca tttatcaaac tatggaaatg accgcctagg gttatatacc tttgtgaact 2400
tggtcaactt tgtgcagagc tggaccaacc tgaaattgca aactctgcct ccagtgcaac 2460
tggcccacca gtattttgag ctcttccctg agcagaaaga ccctctatgg cagaatccat 2520
gtgatgataa acgacacaaa gacatctggt ccagagagaa aacttgtgac cacttaccaa 2580
aatttctagt aattggtcca caaaaaacag gaacaactgc actttattta tttcttctta 2640
tgcatccttc aatcatcagc aatctcccta gtccaaagac atttgaggaa gttcagttct 2700
ttaatggcaa caactatcac aaaggaatag actggtatat ggacttttte cccatccctt 2760
ccaacaccac ctccgacttc tactttgaga aaagcgccaa ctactttgat tcagaagtgg 2820
cgccccggcg ggcagcagcc ctcttgccca aagccaaggt cctgaccatc ctcatcaacc 2880
ccgcggaccg ggcctattcc tggtaccagc accagcgagc ccatgacgac ccagtggccc 2940
taaagtacac cttccatgag gtgattaccg ccggctctga cgcatcctcg aagctgcgtg 3000
ccctccagaa ccgctgcctg gtccctggct ggtacgccac ccacatcgag cgctggctca 3060
gtgcctatca cgccaaccag gtactggtct tggatggcaa actgcttcgc acagaacctg 3120
ccaaagtgat ggacatggtg cagaagttcc ttggggtgac caacaccatt gactaccaca 3180
aaaccttggc gtttgatcca aagaaaggat tttggtgcca actgcttgaa ggaggaaaaa 3240
ccaagtgtct gggcaaaagc aagggccgga aatatcccga gatggacttg cagtcccgag 3300
ccttcctgaa ggactattac cgggaccaca acatcgagct ctccaagctg ctgtataaga 3360
tgggccagac acttcccact tggttacgag aggacttcca gaacaccagg tagccgtgg 3419

<210> 17
<211> 2745
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2860635CB1

<400> 17
atgtgggggc gcacggcgcg gcggcgctgc ccgcgggaac tgcggcgcgg ccgggaggcg 60
ctgttggtgc tcctggcgct actggcgttg gccgggctgg gctcggtgct gcgggcgcag 120
cgtggggccg gggccggggc tgccgagccg ggacccccgc gcaccccgcg ccccgggcgg 180
cgcgagccgg tcatgccgcg gccgccggtg ccggcgaacg cgctgggcgc gcggggcgag 240
gcggtgcggc tgcagctgca gggcgaggag ctgcggctgc aggaggagag cgtgcggctg 300
caccagatta acatctacct cagcgaccgc atctcactgc accgccgcct gcccgagcgc 360
tggaacccgc tgtgcaaaga gaagaaatat gattatgata atttgcccag gacatctgtt 420
atcatagcat tttataatga agcctggtca actctccttc ggacagttta cagtgtcctt 480
gagacatccc cggatatcct gctagaagaa gtgatccttg tagatgacta cagtgataga 540
gagcacctga aggagcgctt ggccaatgag ctttcgggac tgcccaaggt gcgcctgatc 600
cgcgccaaca agagagaggg cctggtgcga gcccggctgc tgggggcgtc tgcggcgagg 660
ggcgatgttc tgaccttcct ggactgtcac tgtgagtgcc acgaagggtg gctggagccg 720
ctgctgcaga ggatccatga agaggagtcg gcagtggtgt gcccggtgat tgatgtgatc 780
gactggaaca ccttcgaata cctggggaac tccggggagç cccagatcgg cggtttcgac 840
tggaggctgg tgttcacgtg gcacacagtt cctgagaggg agaggatacg gatgcaatcc 900
cccgtcgatg tcatcaggtc tccaacaatg gctggtgggc tgtttgctgt gagtaagaaa 960

tattttgaat atctggggtc ttatgataca ggaatggaag tttggggagg agaaaacctc 1020
gaatttttct ttaggatctg gcagtgtggt ggggttctgg aaacacaccc atgttcccat 1080
gttggccatg ttttccccaa gcaagctccc tactcccgca acaaggctct ggccaacagt 1140
gttcgtgcag ctgaagtatg gatggatgaa tttaaagagc tctactacca tcgcaacccc 1200
cgtgcccgct tggaaccttt tggggatgtg acagagagga agcagctccg ggacaagctc 1260
cagtgtaaag acttcaagtg gttcttggag actgtgtatc cagaactgca tgtgcctgag 1320
gacaggcctg gcttcttcgg gatgctccag aacaaaggac taacagacta ctgctttgac 1380
tataaccctc ccgatgaaaa ccagattgtg ggacaccagg tcattctgta cctctgtcat 1440
gggatgggcc agaatcagtt tttcgagtac acgtcccaga aagaaatacg ctataacacc 1500
caccagcctg agggctgcat tgctgtggaa gcaggaatgg ataccttat catgcatctc 1560
tgcgaagaaa ctgccccaga gaatcagaag ttcatcttgc aggaggatgg atctttattt 1620
cacgaacagt ccaagaaatg tgtccaggct gcgaggaagg agtcgagtga cagtttcgtt 1680
ccactcttac gagactgcac caactcggat catcagaaat ggttcttcaa agagcgcatg 1740
ttatgaagcc tcgtgtatca aggagcccat cgaaggagac tgtggagcca ggactctgcc 1800
caacaaagac ttagctaagc agtgaccaga acccaccaaa aactaggctg cattgctttg 1860
aagaggcaat cattttgcca tttgtgaaag ttgtgttgga tttagtaaaa atgtgaataa 1920
gctttgtact tattttgaga acttttttaaa tgttccaaaa taccctattt tcaaagggta 1980
atcgtaagat gttaaccctt ggtatttaga aaattaaaac cttataatat ttttctatca 2040
agatgtatat tttacagtcg tgccttttac tctcattagc aaaaaagata aagattttat 2100
tttggtattt acaagaattc ccaggtacga agatatctgc atgggtggaa atcaggttca 2160
agcaacgtac tttgcattaa ctgataatac ctcagctgcg gggttaaagt tttcccagta 2220
tagagagact gtcactagga acattgtatt gatttattca ggtcattgag atcttctaga 2280
tgtattttaa aaagaatgct ttttggttat gtgttgctac cacagttaac actccataat 2340
gttcatgtca gccaaagagg actaaccaaa gctgaaatct cagagaacaa tttgctttac 2400
taagctgagt caacttgaga gcgaacttct aacaatgccg cactgtagtg tggctggttc 2460
taccactatg actttaaaac atgtttatat catttttaat ttttatgata cggtagtgtc 2520
agggagaaat gtaatgttct atatgaaatt ccttttttcaa gtttgttcat taataacagt 2580
tattaattta aatcagcgtt agagtttgtg ctgctgcaac tgctgtgaaa atttctctga 2640
gtaattctga tttgtgaatg atcccagacc aaccctgaga ttttgtcaac ctgattaagt 2700
caatatgaat gattaaaaag atgtgagaac actgaaaaaa aaaaa 2745

<210> 18
<211> 4763
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2530615CB1

<400> 18
gagagaaata gactcctctt ctgtcccttc taacccaggt ccctacccgt acccctcctc 60
cttcctttcc ccgccccctc ctcccctcctg gggcgagggg ggcctccctc cctctccccc 120
ccttctctct ctccgagggg ggggtcaggg gagggagggg gggtccccca atcagcatgt 180
ggctcctggc gttgtgtctg gtggggctgg ctggggctca acggggagga gggggtcccg 240
gcggcggcgc cccgggcggc cccggcctgg gcctcggcag cctcggcgag gagcgcttcc 300
cggtggtgaa cacggcctac gggcgagtgc gcggtgtgcg gcgcgagctc aacaacgaga 360
tcctgggccc cgtcgtgcag ttcttgggcg tgccctacgc cacgccgccc ctgggcgccc 420
gccgcttcca gccgcctgag gcgcccgcct cgtggcccgg cgtgcgcaac gccaccaccc 480
tgccgcccgc ctgcccgcag aacctgcacg gggcgctgcc cgccatcatg ctgcctgtgt 540
ggttcaccga caacttggag gcggccgcca cctacgtgca gaaccagagc gaggactgcc 600
tgtacctcaa cctctacgtg cccaccgagg acggtccgct cacaaaaaaa cgtgacgagg 660
cgacgctcaa tccgccagac acagatatcc gtgaccctgg gaagaagcct gtgatgctgt 720
ttctccatgg cggctcctac atggagggga ccggaaacat gttcgatggc tcagtcctgg 780
ctgcctatgg caacgtcatt gtagccacgc tcaactaccg tcttggggtg ctcggttttc 840
tcagcaccgg ggaccaggct gcaaaaggca actatgggct cctggaccag atccaggccc 900

tgcgctggct cagtgaaaac atcgcccact ttgggggcga ccccgagcgt atcaccatct 960
ttggttccgg ggcaggggcc tcctgcgtca accttctgat cctctcccac cattcagaag 1020
ggctgttcca gaaggccatc gcccagagtg gcaccgccat ttccagctgg tctgtcaact 1080
accagccgct caagtacacg cggctgctgg cagccaaggt gggctgtgac cgagaggaca 1140
gtgctgaagc tgtggagtgt ctgcgccgga agccctcccg ggagctggtg gaccaggacg 1200
tgcagcctgc ccgctaccac atcgcctttg ggcccgtggt ggatggcgac gtggtccccg 1260
atgaccctga gatcctcatg cagcagggag aattcctcaa ctacgacatg ctcatcggcg 1320
tcaaccaggg agagggcctc aagttcgtgg aggactctgc agagagcgag gacggtgtgt 1380
ctgccagcgc ctttgacttc actgtctcca actttgtgga caacctgtat ggctacccgg 1440
aaggcaagga tgtgcttcgg gagaccatca agtttatgta cacagactgg gccgaccggg 1500
acaatggcga aatgcgccgc aaaaccctgc tggcgctctt tactgaccac caatgggtgg 1560
caccagctgt ggccactgcc aagctgcacg ccgactacca gtctcccgtc tacttttaca 1620
ccttctacca ccactgccag gcggagggcc ggcctgagtg ggcagatgcg gcgcacgggg 1680
atgaactgcc ctatgtcttt ggcgtgccca tggtgggtgc caccgacctc ttcccctgta 1740
acttctccaa gaatgacgtc atgctcagtg ccgtggtcat gacctactgg accaacttcg 1800
ccaagactgg ggaccccaac cagccggtgc cgcaggatac caagttcatc cacaccaagc 1860
ccaatcgctt cgaggaggtg gtgtggagca aattcaacag caaggagaag cagtatctgc 1920
acataggcct gaagccacgc gtgcgtgaca actaccgcgc caacaaggtg gccttctggc 1980
tggagctcgt gccccacctg cacaacctgc acacggagct cttcaccacc accacgcgcc 2040
tgcctcccta cgccacgcgc tggccgcctc gtcccccccgc tggcgccccg ggcacacgcc 2100
ggcccccgcc gcctgccacc ctgcctcccg agcccgagcc cgagcccggc ccaagggcct 2160
atgaccgctt ccccggggac tcacgggact actccacgga gctgagcgtc accgtggccg 2220
tgggtgcctc cctcctcttc ctcaacatcc tggcctttgc tgccctctac tacaagcggg 2280
accggcggca ggagctgcgg tgcaggcggc ttagcccacc tggcggctca ggctctggcg 2340
tgcctggtgg gggccccctg ctccccgccg cgggccgtga gctgccacca gaggaggagc 2400
tggtgtcact gcagctgaag cggggtggtg gcgtcggggc ggaccctgcc gaggctctgc 2460
gccctgcctg cccgcccgac tacaccctgg ccctgcgccg ggcaccggac gatgtgcctc 2520
tcttggcccc cggggccctg accctgctgc ccagtggcct ggggccaccg ccacccccac 2580
cgccccctc ccttcatccc ttcgggccct tccccccgcc ccctcccacc gctaccagcc 2640
acaacaacac gctaccccac ccccactcca ccactcgggt atagggggtg ggtggggagg 2700
ccctcctccc cggccctccc tggcccggcc actccgaagg cagggaggag gacttggcaa 2760
ctggcttttc tcctgtggag tcgtcacacg ccatccagca gcgctaaggt ggacatggga 2820
ttcctccctg cgatgcgtgt ctttcccacg cagagaagcc cagtctcttc tctggatctg 2880
ggcctttgaa caactggggg gcgttttctc ccccccattg ggacaccagt cttcggtgtg 2940
tggaatgtgg tattttcccg cgtggaggtg tgctttctca caacggggtg tgttttccca 3000
tgtgcagggt gaggtttttt tttgccaccc tggacacatg ttggcccccc caaagaattt 3060
ctgtggggat ttgtacccca gaatcctgtt cccccatccc ttctcccacc tcctcccctc 3120
tccctccccc tggagaccct ggaagtggtg tgttcacata cagtgaccct tggccaccag 3180
accacagagg atggagcctg ggaagcagcg aggaaatcac agccccctcg cccctgcctc 3240
ccttgccccc accccggcga agcatgttcc ccccgacgcc ccccttggca caagtcagat 3300
gaagcacgtt ctgccgggga ggccctcacc ttccagagag gacagacaca gatttcctgc 3360
tgggggaggg aggagtccac gcatcctgat gctgcctgga agcttatttt cccgtggcca 3420
ggatgcattt ctctgagtgg aaacaggttc ttgcatgtgg atgtgtgttt ccccaggcag 3480
acggcccctc tcttcccagc acttccctgc ctcccccagg cctcaggccc agcacccagt 3540
tcctcctcac atggcaggtg agcacagact tctagttggc aggagctgag gagggtgaac 3600
aaaccccgag ggaggcccgg cccttgctcc cgagttgggg ggaggggtg tggcaacgtg 3660
ccccccgcag aggccacgca tgtttgacca aagccctcat tgtggtccga ggacagcctt 3720
ttccccaggc ctcagagcat tgctcatccg tgccaaactg ggtaggtgga tttgagcgga 3780
aagactccca aaatgtgcca agaatttccc agtcccaggc agggcagggg aaactaaggg 3840
caagcaggat acagggcgag ggatgtggca ggtgaggggg ctcccgcctg tgccccttct 3900
cctcaccatg tctcccccac cctgcctcag ttctccgttc cccttcatct ccgtcccccc 3960
ctttgaagct gtccccatct cagtgtcaga ccagccttct cctcatctga ccaccctcct 4020
ctgaccgacg ccccctcctt gtctgaaaga aaggagcctt gaatggtgga gggaggcagt 4080
ggggagaaag gtctcaccgg acaggttggg agaatgaggt cagcggtgct ggggaacaga 4140
tggagggggc agtggggaca gggcttgggc agacaccagc aggaataatt tgaaatgtgt 4200
gaggtgactc cccggagggc cttgggcttg ggcatttggg aaaagaatga tgtctggaag 4260

ggcttaaggg acacagtgga cgaggggaga gtcctcatct gctggcattt tgtggggtgt 4320
tagtgccaaa cttgaatagg ggctggggtg ctgtcttcca ctgacaccca aatccagaat 4380
ccctggtctt gagtccccag aactttgcct cttgactgtc ccttctcttc ctacctccat 4440
ccatggaaaa ttagttattt tctgatcctt tcccctgcct ggtctagctc ctctccaaac 4500
agccatgccc tccaaatgct agagacctgg gccctgaacc ctgtagacag atgccctcag 4560
aattggggca tgggaggggg gctgggggac cccatgattc agccacggac tccaatgccc 4620
agctcctctc cccaaaacaa tcccgacaat cccttatccc taccccaacc ctttgcggct 4680
ctgtacacat ttttaaacct ggcaaaagat gaagagaata ttgtaaatat aaaagtttaa 4740
ctgttaaaaa aaaaaaaaaa aaa 4763

<210> 19
<211> 2970
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3883906CB1

<400> 19
cagctaggcg cgagcgaaaa caaacagctg gggctgcgag cgcccccgcc ccggccccga 60
gagcacgccg gcccagtccc ccacctgggg cgcccgtctg cccaccatga ggaagatccg 120
cgccaatgcc atcgccatcc tgaccgtagc ctggatcctg ggcactttct actacttatg 180
gcaggacaac cgagcccacg cagcatcctc cggcggccgg ggcgcgcaga gggcaggcag 240
gaggtcggag cagctccgcg aggaccgcac catcccgctc attgtgacag gaactccctc 300
gaaaggcttt gatgagaagg cctacctgtc ggccaagcag ctgaaggctg gagaggaccc 360
ctacagacag cacgccttca accagctgga gagtgacaag ctgagcccag accggcccat 420
ccgggacacc cgccattaca gctgcccatc tgtgtcctac tcctcggacc tgccagccac 480
cagcgtcatc atcaccttcc acaatgaggc ccgttccacc ctgctgcgca cagtgaagag 540
tgtcctgaac cgaactcctg ccaacttgat ccaggagatc attttagtgg atgacttcag 600
ctcagatccg gaagactgtc tactcctgac caggatcccc aaggtcaagt gcctgcgcaa 660
tgatcggcgg gaagggctga tccggtcccg agtgcgtggg gcggacgtgg ctgcagctac 720
cgttctcacc tttctggata gccactgcga agtgaacacc gagtggctgc cgcccatgct 780
gcagcgggtg aaggaggacc acacccgcgt ggtgagtccc atcattgatg tcatcagtct 840
ggataatttt gcctaccttg cagcatctgc tgaccttcgt ggagggttcg actggagcct 900
gcatttcaag tgggagcaga tccctcttga gcagaagatg acccggacag accccaccag 960
gcccataagg acgcctgtca tagctggagg aatcttcgtg atcgacaagt cctggtttaa 1020
ccacttggga aagtatgatg cccagatgga catctggggg ggagagaatt ttgagctctc 1080
cttcagggtg tggatgtgtg gtggcagtct ggagatcgtc ccctgcagcc gggtgggcca 1140
tgtcttcagg aaacggcacc cctacaactt ccctgagggt aatgccctca cctacatcag 1200
gaatactaag cgcactgcag aagtgtggat ggatgaatac aagcaatact actatgaggc 1260
ccggccctcg gccatcggga aggccttcgg cagtgtggct acgcggatag agcagaggaa 1320
gaagatgaac tgcaagtcct tccgctggta cctggagaac gtctacccag agctcacggt 1380
ccccgtgaag gaagcactcc ccggcatcat taagcagggg gtgaactgct tagaatctca 1440
gggccagaac acagctggtg acttcctgct tggaatgggg atctgcagag ggtctgccaa 1500
gaacccgcag cccgcccagg catggctgtt cagtgaccac ctcatccagc agcaggggaa 1560
gtgcctggct gccacctcca ccttaatgtc ctcccctgga tccccagtca tactgcagat 1620
gtgcaaccct agagaaggca agcagaaatg gaggagaaaa ggatctttca tccagcattc 1680
agtcagtggc ctctgcctgg agacaaagcc tgcccagctg gtgaccagca agtgtcaggc 1740
tgacgcccag gcccagcagt ggcagctgtt gccacacaca tgacggaggg attttcttga 1800
agggtcagct attgcactgt gctgggaggc tggctgtggc tcctgtttgt gaagagaact 1860
cccagttcct tttcacagcc gctgagaaca catccacaca tctaccccac ggccttgtct 1920
tggaactgct cctgctatcc ccacaccctg cttcctcacc ccactggctc ttgggcaaat 1980
gactgttggt accaggaccc tgggggtctc ccctacaaag cagaccagcc tggggcagat 2040
atacctacca caggagcccc ctgtctttca taggccaagg atccacatac ccatagagtt 2100
catggattct agagggtcta ggaatctctt gaaattgtcc ctaaaatact gcatgtgtgt 2160

gcatacatgc attttcctgg ggaaaaagtc catggcttct aaaagaggtc cctgatcccc 2220
aaagggcctg aacctctgct ctggattgaa gccactgtct gcccaagctg ccacccccta 2280
atcttcctcc ctggcgtgct caaactgcca tcgcctgctc cctccacacg gcccttgggg 2340
tcagcagcca agtgtctgtg gtccgcagca cctgctctgg agggctctcc agcagtttcg 2400
cctcctgact ctcaccagcg tcctctcggc agcactcccg tgcccaagtg cacccctctg 2460
gacttgccag cgcaggcccc ttcgctccca gggccatgct ttgcctgtcc tctgtcgtga 2520
tgtttcttcc gcagccaggt gcagcctcag atcccctgtt catctgggaa gcctcctgcc 2580
acagcttagg acagaactgg gcccagggca aaggcctctc cccagaggat ggactagaag 2640
gcctgggcca cactcgggca tggacctttg gggctgggga gcaggggctg cgcctgttga 2700
atgtaagagg actgctgacc agagggcctt caagagggtc tctctgtcct ttgctgtggt 2760
cagatcaggc tctgcactta tcagccggtc ctttgtggca acgcagccct gttctgtttt 2820
tgcttttcct cttcttgacc aaagcatgtg ccactagctg tccttgagga cctcgtcttt 2880
atgaaacaca cacctggaat aaaaccactt cttacatgtc caaaaaaaaa aaaaaaaaaa 2940
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2970

<210> 20
<211> 834
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473644CB1

<400> 20
atgtcgtcct gcagccgcgt ggccctggta actggggcta acaaaggcat cggctttgcg 60
atcacgcgtg acctgtgtcg gaaattctcc ggggacgtgg tgctcacggc gcgggacgag 120
gcgcggggcc gcgcggcggt gcagcagctg caggcggagg gcctgagccc acgcttccac 180
cagctggaca tcgacgaccc gcagagcatc cgtgcgctgc gcgactttct gcgcaaggag 240
tacgggggac ttaacgtgct ggtcaacaac gcgggcatcg cctttagaat ggatgaccca 300
acacccttcg acattcaagc tgaggtgaca ctgaagacga acttttttgc cactagaaat 360
gtctgcactg agttactgcc tataatgaaa ccacatggca gagtggtgaa tatatcaagc 420
ctaataagtc tagaggccct gaaaaactgc agcctggagc tacagcagaa gtttcgaagt 480
gagaccatca cagaggagga gctggtgggg ctcatgaaca agtttgtgga ggatacaaag 540
aaaggagtcc atgcaaaaga aggctggcct aatagtgcat acggggtgtc taagattgga 600
gtgacagtcc tgtccagaat ccttgccagg aaactcaatg agcagaggag aggggacaag 660
atccttctga atgcctgctg ccctggctgg gtcagaaccg acatggcagg accacaagcc 720
accaaaagcc cagaagaagg agcagagacc cctgtgtact tggccctttt gcctccagat 780
gcagagggac ctcatgggca gtttgttcaa gataaaaaag tggaacaatg gtga 834

<210> 21
<211> 2466
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485303CB1

<400> 21
atggtaaatg gcttgtttct cttctccctc tacacacccc aggccccgtc tatgagctct 60
aaaaacaacc ttgaagcctt cagaaggcct gagcagcgag gccaccgggc cgcgcgctcc 120
cagcttcgct cggacgcggc ttcggcccgc agagggttcg tggcccggac gcggcgagag 180
ctgggcccag gacggtgcgt ccggcctcgc ccgcggctgc tcgcaccaac aagtttgaac 240
aatgatcacc gtcaaccccg atgggaagat aatggtcaga agatgcctgg tcaccctgag 300
acccctttcgg ctttttgtcc tgggcatcgg cttcttcact ctctgcttcc tgatgacgtc 360

tctgggaggc cagttctcgg cccggcgcct gggggactcg ccattcacca tccgcacaga 420
aggatgcccc ctggggccgg cctcatggag cggatccagg ctattgccca gaacgtctcc 480
gacatcgctg tgaaggtgga ccagatcctg cgccacagtc tgctcctgca cagcaaggtg 540
tcagaaggcc ggcgggacca gtgtgaggca cccagtgacc ccaagttccc tgactgctca 600
gggaaggtgg agtggatgcg tgcccgctgg acctctgacc cctgctacgc cttctttggg 660
gtggacggca ccgagtgctc cttcctcatc tacctcagtg aggtcgagtg gttctgcccc 720
ccgctgccct ggaggaacca gacggctgcc cagagggcac ccaagccccc ccccaaagtc 780
caggcagttt tccgaagcaa cctgtcccac cttctggacc tgatgggcag cgggaaggag 840
tccctgatct tcatgaagaa gcggaccaag aggctcacag cccagtgggc gctggctgcc 900
cagcgcctgg cacagaagct gggggccacc cagagggacc agaagcagat cctggtccac 960
atcggcttcc tgacggagga gtccggggac gtgttcagcc ctcgggtcct gaagggcggg 1020
cccctagggg agatggtgca gtgggcggac attctgactg cactctatgt cctgggccat 1080
ggcctgcggg tcacagtctc cctgaaggag ctgcagagta acttaggggt accgccaggc 1140
cggggaagct gcccgctcac catgcccctg cccttcgacc tcatctacac cgactaccac 1200
ggcctgcagc agatgaagcg gcacatggga ctctccttca agaagtaccg gtgccgaatc 1260
agggtcatcg acaccttcgg gacggaacct gcgtacaacc acgaggagta cgccacgctg 1320
cacggctacc ggaccaactg gggctactgg aacctcaacc ccaagcagtt catgaccatg 1380
tttcctcata ccccccgacaa ctccttcatg ggcttcgtgt ccgaggagct caacgagacg 1440
gagaagcggc tcatcaaagg cggcaaggcc agcaacatgg ccgtggtgta cggcaaggag 1500
gcgagcatct ggaaggggaa ggagaagttc ctgggcatcc tgaacaaata catggagatc 1560
catggcaccg tgtactacga gagccagcgg ccccccgagg tgccagcctt tgtgaagaac 1620
cacggcctct taccgcagcc tgagtttcag cagctgctgc gcaaggccaa actcttcatc 1680
gggtttggct tcccctacga gggccccgcc cccctggagg ccatcgccaa tggttgcatc 1740
ttcctgcagt cccgcttcag cccgccccac agctccctca accacgagtt cttccgaggc 1800
aagcccacct ccagagaggt gttctcccag catccctacg cggagaactt catcggcaag 1860
ccccacgtgt ggacagtcga ctacaacaac tcagaggagt ttgaagcagc catcaaggcc 1920
attatgagaa ctcaggtaga cccctaccta ccctacgagt acacctgcga ggggatgctg 1980
gagcggatcc acgcctacat ccagcaccag gacttctgca gagctccaga ccctgcccta 2040
ccagaggccc acgccccgca gagccccttt gtcctggccc ccaatgccac ccacctcgag 2100
tgggctcgga acaccagctt ggctcctggg gcctggcccc ccgcgcacgc cctgcgggcc 2160
tggctggccg tgcctgggag ggcctgcacc gacacctgcc tggaccacgg gctaatctgt 2220
gagccctcct tcttcccctt cctgaacagc caggacgcct tcctcaagct gcaggtgccc 2280
tgtgacagca ccgagtcgga gatgaaccac ctgtacccgg cgttcgccca gcctggccag 2340
gagtgctacc tgcagaagga gcctctgctc ttcagctgcg ccggctccaa caccaagtac 2400
cgccggctct gcccctgccg cgacttccgc aagggccagg tggccttgtg ccagggctgt 2460
ctgtga 2466

<210> 22
<211> 1444
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6970969CB1

<400> 22
aacagtgctg ccaagtttgg cctaaggtag cgactggaac gaagacagcc ccccgaatac 60
gcagccgccc agtccgacca tcaggatcgc gaacgccgcc ttggaagaca ggacgcggcg 120
gcttgcgcgc caggaaatcc tggcggttgg cacattcgcg tgctgtcgca accgagtgcc 180
catgcgccag aacatcacca ccccggcgag gctggcgagc ccggcgatat agaacgcggt 240
tgtcaggggc aggtccagcg cagtggcggc gcgtcccagc aaagggccgg aacctatgcc 300
gctcaacata ctgcccgaga gcaaggcgaa atatttggcc cgctggcgcg gttccaccag 360
cagcgtcacg atgatcgggc ccagggtata aaacacgccc cagcccaaac caagagacaa 420
tccaaacagc atcaagcctt ggccaaaccc tggcgtcagg gcaaacccca ggcaggccac 480
caccaacaat acgccggtgc ttgcaatggc ccgcgccgcg cccagcgcgt cagccagatg 540

cccggaaagc acgaccgcta caaaggtgct gagcatcgcc gtggcaatga cactgcccgc 600
gtcgtgttca ttaccgccac gggcatggat cagcagtgac aacaaaaagg tcgagccgta 660
agacagcgac aacagaaaac tggcgaggca gaacagcgcg aatgtcttcg tgcccacggg 720
tggcgtagcg tgcatgagca ggtctctgat cgttggattt atgctcattg tttaccatgg 780
gcgtgcgcag gctttgttct gtggttgctg atgcttgcat caccccgcgc cggagtgaaa 840
cgcgctcgtc ggcatcaatt gcactgcaaa cggccatcaa accgtcttag tatgtggcct 900
ttgaatttgc ccaaggctcg cccatgacca tcgagattcg ccccgccgtg cccagcgatg 960
ctgcgcagat cctgaccttt atcactgaac ttgccgagta cgagaaagct cggcatgaag 1020
tgatcgccag cgtggtggac atcgagcgca gcctcttcag cgaaggcgcc acggcccatg 1080
gcctgatctg cttgcgcgac ggcttgccga ttggctttgc ggtgttcttt ttcagttact 1140
cgacgtggct gggcagtaac tgcctgtacc tcgaagacct gtacatcaac ccggagcaac 1200
gcggcggcgg ggcgggcaag aagttgttgc gccacttggc caagatcgcc tttgacaatg 1260
gttgcgggcg ctttgaatgg agcgtgctgg actggaacga gccagcgatt gccttctaca 1320
aatccatcgg cgcccagcct caggaagaat gggtacgtta tcgcatggaa ggcgatgccc 1380
tgcgcgactt cgccctgggc tgaaacctct ttcgtatagt cactaaagat cgaatgtggg 1440
agcg 1444

<210> 23
<211> 2232
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6538080CB1

<400> 23
atggtggtgg tgcctttcac aattccattc gattcatcag tctggctgct gcagaagttg 60
aacagccccc ggagaatgac tgtagactac cacacactca accaagcggt aactctgatc 120
atagctgctg tgctaggtgt ggcatccctg cttgagcagg gtttcagcag agcaatctgg 180
gagaacaggg agagcctatc aaaggtctgt gtctcccagg ggccatccag gccactggcc 240
tgtgctacta atggggacat caaggtccaa ggagggcctt ctgctgaagg gccacagagg 300
aacaccaggc tgggatggat tcagggcaag caagtcactg tgctgggaag ccctgtgcct 360
gtgaacgtgt tcctcggagt ccccctttgct gctcccccgc tgggatccct gcgatttacg 420
aacccgcagc ctgcatcgcc ctgggataac ttgcgagaag ccacctccta ccctaatttg 480
tgcctccaga actcagagtg gctgctctta gatcaacaca tgctcaaggt gcattacccg 540
aaattcggag tgtcagaaga ctgcctctac ctgaacatct atgcgcctgc ccacgccgat 600
acaggctcca agctccccgt cttggtgtgg ttcccaggag gtgccttcaa gactggctca 660
gcctccatct ttgatgggtc cgccctggct gcctatgagg acgtgctggt tgtggtcgtc 720
cagtaccggc taggaatatt tggtttcttc accacatggg atcagcatgc tccggggaac 780
tgggccttca aggaccaggt ggctgctctg tcctgggtcc agaagaacat cgagttcttc 840
ggtggggacc ccagctctgt gaccatcttt ggcgagtccg cgggagccat aagtgtttct 900
agtcttatac tgtctcccat ggccaaaggc ttattccaca aagccatcat ggagagtggg 960
gtggccatca tcccttacct ggaggcccat gattatgaga agagtgagga cctgcaggtg 1020
gttgcacatt tctgtggtaa caatgcgtca gactctgagg ccctgctgag gtgcctgagg 1080
acaaaaccct ccaaggagct gctgaccctc agccagaaaa caaagtcttt cactcgagtg 1140
gttgatggtg ctttctttcc taatgagcct ctagatctat tgtctcagaa agcatttaaa 1200
gcaattcctt ccatcatcgg agtcaataac cacgagtgtg gcttcctgct gcctatgaag 1260
gaggctcctg aggtcctcag tggctccaac aagtcccttg ccctccatct gatacaaaac 1320
atcctgcaca tcccgcctca gtatttgcac cttgtggcta atgaatactt ccatgacaag 1380
cactccctga ctgaaatccg agacagtctt ctggacttgc ttggagatgt gttctttgtg 1440
gtccctgcac tgatcacagc tcgatatcac agagatgctg gtgcacctgt ctacttctat 1500
gagtttcggc accggcctca gtgctttgaa gacacgaagc cggcttttgt caaagccgac 1560
cacgctgatg aagtccgctt tgtgttcggt ggtgccttcc tgaaggggga cattgttatg 1620
ttcgaaggag ccacggagga ggagaagtta ctgagccgga agatgatgaa atactgggct 1680
acctttgctc gaaccgggaa tcctaatggg aacgacctgt ttctgtggcc agcttataat 1740

ctgactgagc agtacctcca gctggacttg aacatgagcc tcggacagag actcaaagaa 1800
ccgcgggtgg agttttggac cagcaccatc cccctgatcc tgtctgcctc cgacatgctc 1860
cacagtcctc tttcttcctt aactttcctc tctctcctcc agcctttctt tttcttttgt 1920
gctccttgag aagttatctt tctgtgattt tggtttccct tctcctccca taatttctcc 1980
cgcaatcatt agcttctttc tgagctcagc tgctttctat ggggatcctt gcaaaacaag 2040
ctgctttcgc tgatattttа tggacttagg aatgatcctt acagaattct tttcaacatc 2100
aaaaagtgca atttgtcttg gaaggcaaca agatttcttc aataaatttg gaagagggct 2160
ggcctattag ttgtcataat aatggttttg taactcatat gaaataaaat cagaatgtaa 2220
aataggtaaa aa 2232

<210> 24
<211> 816
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55048919CB1

<400> 24
tggatagcgg ctttgacgat cctgggactc gcaacggcaa ccgcggcgat tcgcgcaaac 60
ttccgcgccg aaagccgtca agagcaacat cgttcaggcc ggcggcggca agatcggaga 120
ggtgaccttg accaacggtc gcagtggcgt gcttctgcgt ctgacggcgt ccgggctgac 180
gcctggctgg cacgccatgc atttccatgc gactggcgat tgctccgacc aaggcttcca 240
gaaatcgggc gcgcatatca accacgagga tcacaagaca cctcacgggc ttctcaatcc 300
ggaagggccg gactttgggg aattgccgaa tattcatgtg gcggcggacg gcaccgtcaa 360
tgccgaagcc ttctcggcac ttgtctcgct ggacgcagcg tcatcccggc cgaaccttct 420
cgatgccgat gggtccgccc tggttattca cgccagcccc gatgaccatg tcacgcaacc 480
gatcggcggg gcgggcgcac gcgttgcgtg tgccgttatc cggtaagcgc cagcataccc 540
agcacgggac gcgtgcggtc cgccgatcga ccgcatggcg gatttttaac gaattcggcg 600
aggtgcgcct tagccacgcc gtgttcggcg cgaccaatgg atggagagaa taacgcgttt 660
acgcgccgca aaagcggcga cccagaaaat cagcaacttt ccggcccgca acttgagtga 720
aatttagaag gtaggataat atgatcataa tcaacccgcg ctcagcttga tatcttccgg 780
cgcagatcaa ccagtaagtg gagaagatga cgacat 816

<210> 25
<211> 922
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485135CB1

<400> 25
atggcagaga agcccaagct ccactactcc aatgcacagg gcagaatgga gtccacctgg 60
tggctcctgg ctgcagccgg agtagagttt gaagagaaat ttctagaatc tgcagaagat 120
ttggaaaagt taagaaatga tgggagtttt atgttccaac aagtgccaat ggttgagatt 180
gatgggatga agctggttca gaccagagcc attctcaact acgttgccag caaatacaat 240
ctctatcgga aagacataaa ggagagagtc ctgattgata tgtacacaga aggtatagca 300
gatttgggtg aaatgatcct tcctctgccc atatgtcaac ctgaggataa agacaccaag 360
cttgccttga accaacagag aacaaaaaat cgctatttcc ctgcctttga aaaagtctta 420
aagagcaatg gacaagacta ccttgttggc aacaagctga gctgggctga cattcacttg 480
gtggaacttc tctactacat ggaagagctt gactctagcc ttatcttcag cttccctctg 540
ctgaaggccc tgaaaaccag aatcagcaac ctgcccatgg tgaagaaatt tctacagcct 600
ggcagcccaa ggaagtctct catggatgag aaatctttag aagaagcaag gaagattttc 660

aggttttaat aacgcagtca tggaggccaa gaacttgcaa taccaatgtt ctaaagtttg 720
caacaataaa gtactttacc taagtgttga ttgtgcctgt tgtgaagcta atgaactctt 780
tcaaattata tgctaattaa ataatacaac tcctattcgc tgacttagtt aaaattgatt 840
gttttcatta ggatctgatg tgaattcaga tttccaatct tctcctagcc aaccattttc 900
ctggaattaa aaattcagta aa 922

<210> 26
<211> 1084
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7684978CB1

<400> 26
atcggccgca tggtcgccag tggcctgggt gtgagcgcgg tgccggcgct gtgcatcaac 60
cagatgcagg aattgggcgc ccgctgcgtg acgctggtcg aacccgccgt cgaacggcgt 120
atcggcgtga ttgccctggc cgatcacaaa ctttccaccg cagcccaggc gctgctcgag 180
gtactgcttt cccataccca gactccggag gtgacatgcg ttacgtgaaa ctggcaggtt 240
gcagcgtgcc ggccattggc caaggcaccc ggtacatggg cgaagacccg gcgcgcaagg 300
ccgccgaagt ggccgccctg caatacggta tcgaacaggg catgaccctg atcgacagtg 360
ccgaaatgta tgccgaaggc ggcgccgaga aagtggtcgg ccaagccatt gccgggcgcc 420
gtgatcaggt gtttctggtt agcaaggtgt acccgcacaa cgccagccgc cggggtgtgc 480
ctgaagcctg cgaacgcagc ctgaagcgcc tgggtaccga ggtgatcgac ctttatctgc 540
tgcactggcg tggccagtat ccgctggagg aaacggtcga ggcgttcgag cgcttgcgcg 600
agcagggcaa gattcgccac tgggggtat ccaacttcga tgtcgatgac ctgcgcgaac 660
tgcacaaccc ggattgcgcc accaatcagg tgctgtacaa cccggcgcag cggggcattg 720
aattcgacct gttgccgtgg agccgccaac ggggcttgcc gaccatggcc tactgcccgc 780
tggcccaggc tggccagctg ttgcagcacc cggtgctgag cgagattgcc gaacgccatg 840
gcgcaaggcc ggcgcaggtc agcctggcct gggtgacacg ggacaatgat gtgattgcca 900
ttcccaaggc ggtgtcatca gaacacgtgc ggctgaatgc ggctgccggg gcgttgaccc 960
tgacaagtga agacttgcgg gcgatcgacc gggcgtttcc cgcaccaacg cgcaaacagc 1020
atttggcgat ggtctgactc aagctcagtg aaagttccgg ctctttacat ccagcccctg 1080
cagc 1084